Peptide Models XXXVI

Relative Stability of Major Types of β-Turns as a Function of Amino Acid Composition: A Study Based on Ab Initio Energetic and Natural Abundance Data**

András Perczel,*[a] Imre Jákli,[a] Michael A. McAllister,[b] and Imre G. Csizmadia[a, c]

Abstract: Folding properties of small globular proteins are determined by their amino acid sequence (primary structure). This holds both for local (secondary structure) and for global conformational features of linear polypeptides and proteins composed from natural amino acid derivatives. It thus provides the rational basis of structure prediction algorithms. The shortest secondary structure element, the β-turn, most typically adopts either a type I or a type II form, depending on the amino acid composition. Herein we investigate the sequence-dependent folding stability of both major types of β-turns using simple dipeptide models (-Xxx-Yyy-). Gas-phase ab initio properties of 16 carefully selected and suitably protected dipeptide models (for example Val-Ser, Ala-Gly, Ser-Ser) were studied. For each backbone fold most probable side-chain conformers were considered. Fully optimized 3-21G RHF molecular structures were employed in medium level [B3LYP/6-311++G(d,p)//RHF/3-21G] energy calculations to estimate relative populations of the different backbone conformers. Our results show that the preference for β-turn forms as calculated by quantum mechanics and observed in X-ray determined proteins correlates significantly.

Keywords: ab initio calculations · β-turns · conformation analysis · protein folding · protein models

Introduction

Studies revealing correlations between conformation and molecular function of the different building units of peptides and proteins have been in the frontier of chemistry and biochemistry. Secondary structural elements can either be composed of homo- or hetero- as well as of repetitive or nonrepetitive conformational subunits. In a homo-conformer the appropriate backbone dihedral value of residue i is close to that of $(i-1)$ or $(i+1)$, for example, $|\xi(i)-\xi(i\pm1)|\leq 15^\circ$ ($\xi=\phi$ or ψ). On the other hand, in hetero-conformers the same parameters are distinctly different between neighboring residues. α- and 3_{10}-helices, β-pleated sheets, and collagen or poly-proline II structures are made of typical homo-conformational subunits. The β-turn is the shortest secondary structural element of globular proteins, containing two central amino acid residues, $i+1$ and $i+2$, embedded in a tetrapeptide sequence unit, labeled from i to $i+3$. In several types of β-turns the adjacent values of ϕ and ψ are significantly different. Thus, β-turns can be typical secondary structural elements of proteins composed from hetero or nonrepetitive conformational building units, $|\xi(i)-\xi(i+1)|\geq 30^\circ$ where $\xi=\phi$ or ψ. A good example is that of a type II β-turn with $\phi(i+1)_{\text{type II}}\approx -60^\circ$, $\psi(i+1)_{\text{type II}}\approx 120^\circ$, $\phi(i+2)_{\text{type II}}\approx 80^\circ$ and $\psi(i+2)_{\text{type II}}\approx 0^\circ$ backbone parameters, where $\phi(i+1)$ differs from $\phi(i+2)$ by 140° and $\psi(i+1)$ from $\psi(i+2)$ by 120°.

β-Turns are involved in the reversal of the main chain (e.g. β-sheet–β-turn–β-sheet motif), thus they directly influence the proper fold of the macromolecule. Sequences adopting a β-turn structure can be the target of several post-translational modifications (glycosylation, phosphorylation) or immune recognition. Furthermore, β-turns may have an important role in the unfolding-refolding process of proteins; the β-turn structure can be retained, even in a partially unfolded or in a

[a] Prof. A. Perczel, I. Jákli, I. G. Csizmadia
Department of Organic Chemistry, Eötvös L. University
1117 Budapest (Hungary)
Fax: (+36) 1-209-0602
E-mail: perczel@para.chem.elte.hu

[b] M. A. McAllister
Agouron Pharmaceutical
San Diego, CA 92037-1020 (USA)

[c] I. G. Csizmadia
Department of Chemistry
University of Toronto
Ontario, ON, M5S 1A1 (Canada)

[**] Peptide Models, Part 36; for Part 35 see P. Hudáky, I. Hudáky, A. Perczel, *J. Mol. Struct.* (*THEOCHEM*) **2002**, *583*, 199–213.

 DOI: 10.1002/chem.200204393

molten globule state. Thus, turns could have a key role in structure restoration of proteins. The hundreds of experimental and theoretical studies associated with β-turns in the last thirty years show that although it is the shortest and simplest secondary structural element of proteins it is of great importance.[1–3]

Different forms of β-turns (Table 1) are most commonly distinguished: I, I′, II, II′, III, III′, VIa, VIb and VIII.[4] In globular proteins the different types of β-turns exhibit very different natural abundances.[5–7]

Table 1. Selected conformational parameters of β-turns.

Type of β-turns	Dihedral angles [°]			
	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	−60	−30	−90	0
I′	60	30	90	0
II	−60	120	80	0
II′	60	−120	−80	0
III	−60	−30	−60	−30
III′	60	30	60	30
IV	types I–III′ turns with two or more torsional angles deviating more than 40° form the ideal values of Venkatachalam			
V	−80	80	80	−80
V′	80	−80	−80	80
VIa	*cis* X-Pro bond, where X is residue $i+1$			
VII	$\phi_2 \approx 180$,		$\lvert\phi_3\rvert < 60$	
	or			
	$\lvert\phi_2\rvert < 60$,		$\phi_3 \approx 180$	
VIII	−60	−30	−120	120

Abstract in Hungarian: *A fehérjéket felépítő aminosavak szekvenciális sorrendje meghatározza az egész molekula konformációs tulajdonságát. Mindez egyaránt igaz az alfa aminosavakból felépülő lineáris peptidek illetve fehérjék másodlagos és harmadlagos szerkezeti elemeire, ami megteremti a szerkezetbecslő eljárások racionális alapját. A β-kanyar szerkezet a fehérjék legrövidebb másodlagos szerkezeti eleme, mely leggyakrabban I vagy II formát ölt. Ez a szerkezeti különbség a felépítő aminosavak típusában, illetve azok sorrendjében rejlik. Cikkünk egyszerű, -Xxx-Yyy-aminosavösszetételű β-kanyar szerkezetek stabilitását vizsgálja, a korábban említett két főtípus esetében, az aminosavszekvencia függvényeként. Tizenhat gondosan kiválasztott összetételű dipeptid (pl. Val-Ser, Ala-Gly, Ser-Ser) gázfázisú ab initio tulajdonságait tanulmányoztuk. Az összes valószínű oldallánc-térállást figyelembe vettük minden egyes eltérő összetételű és gerinckonformációjú szerkezet esetében. A különböző konformerek relatív gyakoriságát, RHF 3-21G elméleti szinten optimált szerkezetekhez tartozó, közepes szintű [B3LYP/6-311++G(d,p)//RHF/3-21G] energiaszámításokra alapoztuk. Eredményeink azt mutatják, hogy a β-kanyar szerkezetek konformációtípusainak relatív stabilitása a számított kvantumkémiai modellrendszeren belül, illetve ugyanezen szerkezeti elemek relatív gyakorisága fehérjékben (röntgenkrisztallográfiai adatokra támaszkodva) szignifikáns korrelációt mutat.*

Type III turns, as part of 3_{10}-helices, are likely to be the most frequently observed hairpin structure. Type I is the second and type II is the third most abundant form of β-turn in globular proteins. The ideal forms of type I and III β-turns differ form each other only by some 30 degrees of torsional angles $\phi(i+2)$ and $\psi(i+2)$, respectively. Therefore, they are often not distinguished from each other and are labeled as type I(III) β-turns. Thus, when the type preference of β-turns is investigated as a function of their amino acid sequence, most commonly only the ratio of type I(III) to type II is determined.[5, 7, 8] The above-mentioned three forms of β-turns (I, II, and III) constitute more than 95% of all β-turns assigned in proteins.[7] The conformational mirror image structures of the major types of β-turns (type I′, type II′, and type III′) and additional forms of β-turns, containing for example a *cis* amide bond (type VI β-turn), have low natural abundance in proteins.

When removed form their natural environment and investigated in the form of a peptide in solution, amino acid sequences that form typical secondary structural elements in proteins unfold and often adopt multiple conformers with little or no resemblance to their original conformation. Furthermore, these equilibrium structures are in fast exchange, making analyzes at the atomic level complicated. Removed from their natural environment β-turns may show atypical and unexpected conformational features. Nowadays, studying the conformational properties of secondary structural elements, especially turns, by simply using synthetic linear peptides of natural amino acid residues is less common. Several laboratories have made significant contributions by using model systems of restricted motion (e.g. cyclic or bridged peptides).[9–12] Performing IR and NMR studies of these model peptides with computational work at different level of theory has revealed important scientific details.[13–16] Based on primary sequence information, a lot is known about where to locate β-turns in proteins and how to predict their most probable form. However, some issues are still not yet fully revealed. For example, the role of the amino acid side chains in the residues that make up a β-turn is not understood in full detail. Furthermore, it is not fully clear to what extent the adjacent or spatially close residues of a β-turn are involved in the fine-tuning of the secondary structure. What are the explicit effects of the solvent and the molecular environment in stabilizing β-turn structures? Do β-turns form similarly in a hydrophobic region as on the surface of a protein, where it is hydrated? For many reasons, the solution-state conformation analysis of linear di-, tri-, and tetrapeptides was only of limited success.[17, 18] In addition to spectroscopy, more and more accurate theoretical calculations and conformation-dependent stability studies now offer an alternative scientific approach when such questions are addressed.

From the early study of Vancatachalam[19] to the recent book of Sapse[20] theoretical publications have investigated different aspects of β-turns. The computing power of the seventies and eighties made possible the determination of structures of shorter peptides (e.g. β-turns) by means of molecular mechanic (MM) and semiempirical techniques.[21] Since the early nineties more and more ab initio studies on

turns have been published.[22–27] In one of the first ab initio studies of β-turns -Gly-Gly-[22] and -Ala-Ala-[23] model sequences were computed at the 3-21G RHF level of theory. The geometries of three hairpin-forming dipeptides (N- and C-unprotected H-Pro-Ala-OH, H-Pro-D-Ala-OH, and H-Pro-Gly-OH) were also computed at the 6-31G RHF level, at which a *cis* peptide bond was also considered.[26] Möhle et al.[28] conducted a systematic study, in which Aib (α-aminoisobutyric acid), in addition to three natural alpha amino acids (Ala, Gly and Pro), was incorporated into specific locations of the model peptide. They found that the -Ala-Gly- sequence prefers type I while -Gly-Gly- favors type II hairpin conformers, which agrees with most common expectations and with results of sequence predictions based on proteins. The relative stabilities of some of these hairpin conformers (e.g. -Ala-Ala-) with other key structures were also computed at the MP2 level of theory[28–30] in order to explore the effect of electron correlation. The comprehensive analysis of inverse γ-turn (γ_L) and extended (β_L) backbone conformers revealed a rather similar conclusion.[29, 31] The pair-wise comparison of the geometric properties computed at RHF/6-31G(d) and RHF/3-21G levels of theory revealed that backbone torsional values are similar for the different turns.[28]

The clear advantage of any computational method is that all minima on the potential energy surface can be determined, or in other words all relevant structures, even those with exceptionally short lifetimes, can be investigated. For the investigation of conformational libraries, composed of inherently flexible molecules, in which several conformers have low stability (very short lifetimes) a computational approach seems adequate. The first conformational library of this kind determined, which contained the ab initio structures of the For-L-Ala-L-Ala-NH_2 model system, had 49 elements, all with different backbone folds.[23] It was subsequently recalculated by Yu et al.[27] and was completed with two additional conformers. The comprehensive analysis of these 51 structures revealed that more than 60% of these structures could be classified as β-turns using the previously established selection criteria of τ- and d-values.[23] The value of the "virtual" dihedral angle τ ($-180 \le \tau\{C^\alpha(i) - C^\alpha(i+1) - C^\alpha(i+2) - C^\alpha(i+3)\} \le +180^\circ$) measures the openness of a backbone fold, while d is the distance between the α-carbons (or their substituent) in residues i and $(i+3)$. Structures with $|\tau| \le 90^\circ$ and with $d(C^\alpha(i) - C^\alpha(i+3)) \le 7$ Å can be classified as β-turns. Note that most "classical" forms of β-turns (e.g. types I, II, and III) have τ values close to zero with a relatively short d value. Conformers with τ larger than 90° or with d greater than 7 Å are found to be partially or fully extended molecular structures.[24, 27] Although the comprehensive analysis of this conformational library revealed several types of β-turn conformers of For-L-Ala-L-Ala-NH_2, the type I(III) and type II β-turns are the most stable and therefore the most important ones. To understand the effect of side-chain-induced backbone stability on β-turns we have decided to perform systematic ab inito calculations of these two basic turn conformers (type I(III) and type II) for model systems having more complex side chains than alanine.

Scope

The purpose of this paper is to compute and compare geometric and energetic properties of suitably selected sets of β-turn models obtained from calculations and from experimental data. Four simple amino acid residues (Gly, Ala, Val, and Ser), which can form both type I and type II β-turns, were employed. The above four amino acid residues were selected based on the following well-known principles of "residue preference" of β-turns:

1) hydrophobic residues (e.g. Pro, Ala, and Val) are suitable for the $(i+1)$ position of both type I(III) and type II β-turns,[32]
2) at the $(i+2)$ position of a type II β-turn glycine is experimentally observed in proteins at least four times as often as any other amino acid residue, and
3) short and polar side chains (e.g. Ser, Asp, Asn) are preferred at the $(i+2)$ position of a type I β-turn.

Using Gly, Ala, Val, and Ser 16, different For-L-Xxx-L-Yyy-NH_2 type triamides can be constructed, positioning all four amino acid residues at both the $(i+1)$ and $(i+2)$ positions of the β-turn structure.

Both type I and type II β-turns (Table 1) were computed for all 16 models. Unlike for glycine and alanine, the backbone torsional angles of all other amino acid residues are influenced by the relative orientation of the side chain. For example, in any backbone conformers, as many as nine different side-chain orientations are expected for serine and three for valine residues, this results in a total of 27 different side-chain rotamers. The ensemble of these conformers after optimization at a given level of theory is called the conformational library of the model peptide associated with a particular backbone structure. Our present goal was to determine as many side-chain conformers as possible for each of the above mentioned 16 peptide models, in which both type I and type II β-turn backbone structures were considered. Thus, a total of 32 ab initio conformational libraries were computed and analyzed.

From previously published data[33–37] we selected the conformational building units needed to compose either type I ($\alpha_L\delta_L$) or type II ($\varepsilon_L\alpha_D$ or $\gamma_L\alpha_D$) β-turn structures. The shorthand notations (-$\alpha_L\delta_L$-, -$\varepsilon_L\alpha_D$-, or -$\gamma_L\alpha_D$- etc.) introduced earlier for dipeptides[23] will be briefly explained below. We have found[34] that even single-point calculations can provide high-quality relative energies employing RHF/3-21G geometries. In the case of β-turn conformational subunits (α_L, δ_L, and γ_L) both ab initio and DFT (B3LYP) single point energies (e.g. ΔE(B3LYP/6-311++G**//RHF/3-21G)) show high correlation with energies obtained by optimization (e.g. ΔE(B3LYP/6-311++G**)). The R^2 value of the For-L-Val-NH_2 model is similarly high (0.9941). For both types of molecule, the correlation coefficient is significantly lower when RHF/3-21G energies are compared with B3LYP/6-311++G** values (e.g. $R^2_{\text{For-L-Val-NH}_2} = 0.7214$ when ΔE(B3LYP/6-311++G**) and ΔE(RHF/3-21G) are correlated.)

Therefore, the following computational scheme was used for the calculations of β-turn conformational libraries:

1) optimization of all structures at the RHF/3-21G level of theory, and

2) calculation of higher level single point computations by using the B3LYP/6-311++G**//RHF/3-21G level of theory.

At present, this protocol seems to provide the best compromise between the opposing requirements of economy (manageable computer time) and accuracy (reliable results). This was the only strategy that seemed feasible and detailed enough to result in useful structural and adequate stability data for as many as 32 conformational libraries incorporating far more than 150 conformers of diaminoacid triamides.

Methods

Nomenclature for backbone and for side-chain conformers: the Ramachandran map $E=E(\phi,\psi)$ can be divided into conformational regions, also called catchment regions, in many ways. For the torsional angle pair ϕ and ψ, multidimensional conformational analysis (MDCA) predicts nine catchment regions, as depicted in Figure 1 using the g^+, a, g^- terminology (Figure 1A) or the "shorthand" notation suggested earlier for protein conformational building units.[44]

Following the IUPAC-IUB recommendations, the *gauche*$^+$ (g^+), *anti* (a), and *gauche*$^-$ (g^-) descriptors were used for notation of the conformers. In order to simplify this four character (e.g. g^+g^+) notation to a two character symbol (e.g. α_D), a shorthand notation for the typical main chain folds was introduced in the early 1900s:[44] $\alpha_L=(g^-,g^-)$, $\alpha_D=(g^+,g^+)$, $\beta_L=(a,a)$, $\gamma_L=(g^-,g^+)$, $\gamma_D=(g^+,g^-)$, $\delta_L=(a,g^+)$, $\delta_D=(a,g^-)$, $\varepsilon_L=(g^-,a)$, and $\varepsilon_D=(g^+,a)$ (Figure 1C). An alternative shorthand notation for the same type of minima was introduced by Karplus.[39] If two amino acids form the model system, -Xxx-Yyy-, the backbone conformation "code" is composed of the variation of the code associated with the first and the second residue, resulting in a total of 81 possible structures ($\gamma_L\gamma_L$, $\beta_L\gamma_L$, $\delta_L\beta_L$, etc.). The shorthand notation for the backbone fold of a type I β-turn is $\alpha_L\delta_L$, while that of a type II β-turn is primarily $\varepsilon_L\alpha_D$. In the present study of over 200 β-turn conformers, a total of 32 libraries were investigated, all associated with either a type I or type II β-turn backbone fold.

For alanine no side-chain conformation needs to be specified. In valine, which contains two geminal C$^\gamma$ carbons (C$^\gamma_A$ and C$^\gamma_B$) with a proton (H$^\beta$) attached to C$^\beta$, three distinct χ_1 rotamers are expected, labeled as 60, 180, or 300°. In contrast, two torsional angles, χ_1 and χ_2, are present in the case of serine; thus a total of nine rotamers are expected. For all side-chain rotamers the variation of *gauche*$^+$ (g^+), *anti* (a), and *gauche*$^-$ (g^-) nomenclature is used, also abbreviated as $+$, a, and $-$.

Ab initio computations: Both RHF/3-21G geometry optimizations and B3LYP/6-311++G**//RHF/3-21G type single-point calculations were performed with the program package Gaussian 98.[40] As mentioned above, both type I ($\alpha_L\delta_L$) and type II ($\varepsilon_L\alpha_D$ or $\gamma_L\alpha_D$) β-turn conformers were computed in the present study. The $\varepsilon_L\alpha_D$ backbone orientation is a type II β-turn, as predicted by Vancatachalam,[19] with a ten-membered hydrogen bond. However, the $\gamma_L\alpha_D$ conformation also shows strong resemblance to a type II β-turn, but with a seven-membered hydrogen bond. Although $\varepsilon_L\alpha_D$ was always our first choice, occasionally both forms of type II β-turn were considered. Table 2 contains conformational properties found for the side-chain orientations of For-L-Xxx-NH$_2$ models at the RHF/3-21G level of theory associated with any of the following backbone structures: α_L, δ_L, α_D, and γ_L (Table 2). None of these amino acid diamides adopt the ε_L (polyproline II) backbone conformation; thus, $\phi=-60°$ and $\psi=120°$ initial values were used. First, the above-mentioned different types of side-chain orientations (Table 2) were used to construct input conformers. For example, in the case of the type I β-turn ($\alpha_L\delta_L$) of the -Ser-Val- model, a total of six structures were expected: three from the different side-chain orientation of Ser having an α_L and two from Val having a δ_L-type subconformation (Table 2). Second, for completeness all additional possible structures (Table 3) were optimized at the RHF/3-21G level, both for type I and type II β-turns. Thus, one may anticipate the existence of certain conformers in

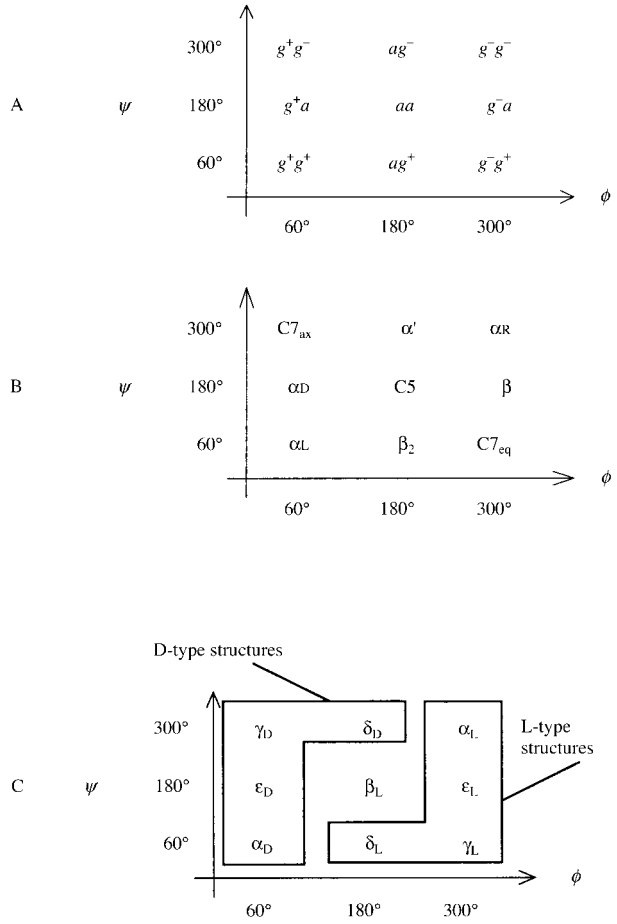


Figure 1. A) The ideal location of the nine basic backbone structures on an $E=E$ (ϕ, ψ) surface, labeled according to the IUPAC-IUB guidance. B) The shorthand notation for the same type of minima used in some other laboratories. C) The abbreviation applied in this paper for the above nine typical backbone conformers using descriptors (α_L, ε_L, γ_L etc.) incorporating most traditional elements of peptide chemistry nomenclature introduced previously (For amino acid residues of L-enantiomeric form basic conformers are more frequent among the set of "L-type structures" than those from the D-"valley").[4] To describe the backbone fold of dipeptide -Xxx-Yyy- any variation of the above nine basic conformers of both Xxx and Yyy is possible, resulting in a maximum of $9^2=81$ ideal structures.

Table 2. Geometric properties of conformational building units of Gly, Ala, Ser and Val amino acid residues in a For-L-Xxx-NH_2 model system required for type I ($\alpha_L\delta_L$) and for type II ($\varepsilon_L\alpha_D$ or $\gamma_L\alpha_D$) β-turns.

	Conformational building units both for type I and type II β-turns							
Xxx	α_L		δ_L		α_D		γ_L	
Gly	no such minima		1	$\phi=-126^\circ$ $\psi=26^\circ$	no such minima		1	$\phi=-84^\circ$ $\psi=68^\circ$
Ala	no such minima		1	$\phi=-128^\circ$ $\psi=30^\circ$	1	$\phi=64^\circ$ $\psi=33^\circ$	1	$\phi=-84^\circ$ $\psi=68^\circ$
Ser[30]	3 (*a*,*a*)[a] (−,*a*) (−,−)	$\phi=-68^\circ$ (5°)[b] $\psi=-30^\circ$ (11°)	4 (−,−) (+,*a*) (−,*a*) (*a*,−)	$\phi=-132^\circ$ (14°) $\psi=30^\circ$ (7°)	4 (+,+) (*a*,+) (−,*a*) (*a*,−)	$\phi=57^\circ$ (7°) $\psi=42^\circ$ (9°)	6 (−,−) (−,*a*) (−,+) (*a*,−) (*a*,+) (+,+)	$\phi=-82^\circ$ (4°) $\psi=67^\circ$ (6°)
Val[31]	no such minima		2 (−)[c] (+)	$\phi=-131^\circ$ (8°) $\psi=32^\circ$ (5°)	3 (−) (*a*) (+)	$\phi=53^\circ$ (7°) $\psi=43^\circ$ (2°)	3 (−) (*a*) (+)	$\phi=-86^\circ$ (1°) $\psi=67^\circ$ (4°)

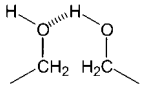
[a] Both for χ^1 and for χ^2 the − (g^-), a (*anti*) and + (g^+) orientations are considered. [b] Average backbone values with standard deviations. [c] Only side-chain torsion χ^1 is considered.

Table 3. Number of input structures and optimized conformers (in parenthesis) at the RHF/3-21G level of theory both for type I(III) and type II β-turn conformers.

Type I(III): $(i+1)/(i+2)$ bb. $=\alpha_L\delta_L$	Gly*(1)	Ala (1)	Ser (9)	Val (3)
Gly (1)[a]	1 (1)[b]	1 (1)	9 (5)	3 (3)
Ala (1)	1 (1)	1 (1)	9 (5)	3 (3)
Ser (9)	9 (7)	9 (7)	25 (24)	27 (17)
Val (3)	3 (3)	3 (3)	19 (15)	9 (9)
Type II: $(i+1)/(i+2)$ bb. $=\varepsilon_L\alpha_D$	**Gly(1)**	**Ala (1)**	**Ser (9)**	**Val (3)**
Gly (1)	1+0 (0+1)[c]	1+0 (1+0)	5+0 (3+2)	3+0 (3+0)
Ala (1)	1+0 (0+1)	1+0 (1+0)	5+0 (5+0)	3+0 (2+1)
Ser (9)	9+0 (1+5)	3+0 (0+3)	18+8 (15+9)	9+2 (7+2)
Val (3)	3+0 (2+1)	3+1 (3+1)	15+0 (15+0)	9+0 (6+3)

[a] Maximum number of theoretically possible side-chain orientations (e.g. 9 for Ser and 81 for -Ser-Ser-). [b] Number of input structures used for RHF/3-21G optimization and number of minima found at this level of theory. [c] For type II β-turns occasionally $\gamma_L\alpha_D$ (or $\gamma_L\delta_L$, $\gamma_L\delta_D$, and $\varepsilon_L\delta_L$) structures (minor forms) were also considered as well as the $\varepsilon_L\alpha_D$ major type of backbone conformation: "major" + "minor" forms are reported. In parenthesis the total number of optimized major + minor form were tabulated [Examples for a few minor forms: $\gamma_L\delta_L$ AG; ($\phi_{i+1}=-76.1^\circ$, $\psi_{i+1}=91.1^\circ$, $\phi_{i+2}=119.2^\circ$ and $\psi_{i+2}=-2.4^\circ$), $\gamma_L\delta_D$ GS; ($\phi_{i+1}=-82.1^\circ$, $\psi_{i+1}=70.9^\circ$, $\phi_{i+2}=-163.9^\circ$ and $\psi_{i+2}=-63.1^\circ$), $\varepsilon_L\delta_L$ SG; ($\phi_{i+1}=-72.8^\circ$, $\psi_{i+1}=109.9^\circ$, $\phi_{i+2}=115.8^\circ$ and $\psi_{i+2}=-12^\circ$)].

triamides (e.g. HCO-Ser-Ser-NH_2), which are not present in simple diamides, provided that there is a special interaction (e.g. hydrogen bond) between the two hydroxymethyl side-chain groups.



For type I β-turn backbone conformations, all possible side-chain rotamers were subjected to minimization, except those of -Val-Ser- and -Ser-Ser-. Table 4 reports conformational and energetic properties computed for 132 type I(III) and for 100 type II β-turn input structures. (Table 4) contains all data of 16 peptide libraries (close to 200 stable and fully optimized molecular structures). Relative populations for computed ab initio relative energies are determined according to $\exp(-\Delta E/RT)/\Sigma\exp(-\Delta E/RT)$, where $RT=NkT=0.595371\ \text{kcal mol}^{-1}$ ($T=300\,\text{K}$, $k=1.38\times10^{-23}\,\text{J K}^{-1}$) and Avogadro's number (N) is $6.02\times10^{23}\ \text{mol}^{-1}$.

Databases: Following the guidance of Hobohm et al., our protein database contains a total of 650 proteins of homology level equal to or lower than 25 %.[41, 42] These proteins were analyzed for sequence unit -Xxx-Yyy- (where Xxx and Yyy = Gly (G), Ala (A), Ser (S), or Val (V)). All entries correspond to high-resolution X-ray structures, no structures determined by NMR spectroscopy were incorporated. Data were taken from the 1998 issue of the Protein DataBase. Both ab initio computed and experimentally determined probabilities of type I(III) and type II β-turns are reported in Tables 5 and 6. Cross-correlation (R^2) values between ab initio computed and experimentally determined probabilities of the ratio of type I(III) and type II β-turns (at a given κ structural tolerance) are reported in Table 7.

Results and Discussion

Structure, computed geometries, and experimental fold: Optimized structures of type I β-turns correspond to the $\alpha_L\delta_L$ backbone conformation of (Table 4). For higher-quality energy values, RHF/3-21G minimizations were followed by B3LYP/6-311++G** single-point computations. Optimization of type II β-turn conformers was achieved, resulting in primarily the $\varepsilon_L\alpha_D$ backbone fold (Table 4). Depending on the amino acid composition (G, A, V, or S) and the side-chain

Table 4. Relative energies [kcal mol^{-1}] and probabilities[a] determined at two levels of theory for all β-turn conformers (level A: RHF/3-21G, level B: B3LYP/6-311++G**//RHF/3-21G).

Model + conformation	Level A		Level B		Model + conformation	Level A		Level B	
Pept. BB. SC. [b]	ΔE[c]	p_i[d]	ΔE	p_i	Pept. BB. SC.	ΔE	p_i	ΔE	p_i
GG aLdL	0.0[e]	**0.521**	0.4	**0.344**	VV aLdL_60–60	1.1	0.076	3.3	0.003
GG gLdL	0.1	**0.479**	0.0[f]	**0.656**	VV aLdL_60–180	2.4	0.008	4.0	0.001
AA aLdL	0.0[g]	**0.900**	0.0[h]	**0.984**	VV aLdL_60–300	0.4	**0.245**	1.3	**0.101**
AA eLaD	1.3	**0.100**	2.4	0.016	VV aLdL_180–60	2.2	0.011	3.9	0.001
SS aLdL_*aa ag*$^-$	13.1	0	8.3	0	VV aLdL_180–180	2.6	0.005	4.4	0.001
SS aLdL_*aa g*$^+$*g*$^+$	11.0	0	7.5	0	VV aLdL_180–300	1.0	0.082	2.0	0.027
SS aLdL_*ag*$^-$ *g*$^-$*a*	9.2	0	8.6	0	VV aLdL_300–60	1.0	0.080	2.0	0.028
SS aLdL_*ag*$^-$ *g*$^-$*g*$^-$	8.0	0	8.5	0	VV aLdL_300–180	1.9	0.020	2.7	0.009
SS aLdL_*ag*$^+$ *g*$^-$*g*$^+$	12.0	0	9.4	0	VV aLdL_300–300	0.0[k]	**0.467**	0.0[l]	**0.828**
SS aLdL_*g*$^-$*a ag*$^-$	5.3	0	1.4	0.021	VV eLaD_60–60	6.1	0	7.5	0
SS aLdL_*g*$^-$*g*$^-$ *ag*$^-$	5.2	0	1.2	0.028	VV gLaD_60–180	4.0	0.001	6.6	0
SS aLdL_*g*$^-$*a g*$^+$*g*$^+$	4.3	0.001	1.3	0.022	VV eLaD_60–300	5.4	0	6.1	0
SS aLdL_*g*$^-$*g*$^-$ *g*$^+$*g*$^+$	4.3	0.001	1.1	0.032	VV eLaD_180–60	5.6	0	6.3	0
SS aLdL_*g*$^-$*a g*$^-$*a*	10.5	0	5.5	0	VV gLaD_180–180	2.8	0.004	4.9	0
SS aLdL_*g*$^-$*a g*$^-$*g*$^-$	8.7	0	4.6	0	VV eLaD_180–300	4.9	0	5.0	0
SS aLdL_*g*$^-$*a g*$^-$*g*$^+$	11.5	0	4.7	0	VV eLaD_300–60	6.0	0	7.1	0
SS aLdL_*g*$^-$*g*$^-$ *g*$^-$*a*	10.4	0	5.3	0	VV gLaD_300–180	4.1	0	6.1	0
SS aLdL_*g*$^-$*g*$^-$ *g*$^-$*g*$^+$	11.4	0	4.5	0	VV eLaD_300–300	5.2	0	5.7	0
SS aLdL_*g*$^-$*g*$^-$ *g*$^-$*g*$^-$	8.9	0	4.5	0	GA aLdL	0.0[m]	**0.978**	0.0[n]	**0.962**
SS aLdL_*g*$^+$*a ag*$^-$	4.6	0	0.5	0.092	GA eLaD	2.3	0.022	1.9	0.038
SS aLdL_*g*$^+$*g*$^+$ *ag*$^-$	4.8	0	0.0	0.202	AG aLdL	1.2	**0.123**	0.9	**0.192**
SS aLdL_*g*$^+$a *g*$^+$*g*$^+$	4.2	0.001	0.4	0.104	AG gLdL	0.0[o]	**0.877**	0.0[p]	**0.808**
SS aLdL_*g*$^+$*g*$^+$ *g*$^+$*g*$^+$	4.3	0.001	0.0	0.207	GS aLaD_*ag*$^-$	1.4	**0.087**	0.2	**0.392**
SS aLdL_*g*$^+$*a g*$^-$*g*$^+$	8.1	0	4.4	0	GS aLdL_*g*$^+$*g*$^+$	0.0[q]	**0.895**	0.0[r]	**0.546**
SS aLdL_*g*$^+$*a g*$^-$*g*$^-$	9.5	0	4.3	0	GS aLaD_*g*$^-$*a*	7.3	0	3.5	0.002
SS aLdL_*g*$^+$*g*$^+$ *g*$^-$a	3.2	0.004	2.7	0.002	GS aLaD_*g*$^-$*g*$^+$	7.3	0	3.4	0.002
SS aLdL_*g*$^+$*g*$^+$ *g*$^-$*g*$^-$	2.3	0.020	2.5	0.003	GS aLaD_*g*$^-$*g*$^-$	5.2	0	3.3	0.002
SS aLdL_*g*$^+$*g*$^+$ *g*$^-$*g*$^+$	10.8	0	3.1	0.001	GS eLaD_*ag*$^-$	2.3	**0.018**	1.4	**0.051**
SS eLaD_*ag*$^-$ *g*$^+$*g*$^+$	5.2	0	5.7	0	GS eLaD_*g*$^-$*a*	7.8	**0**	4.0	**0.001**
SS eLaD_*ag*$^-$ *g*$^-$*a*	7.2	0	5.0	0	GS eLaD_*g*$^-$*g*$^-$	8.8	**0**	4.7	**0**
SS eLaD_*ag*$^+$ *g*$^-$*g*$^+$	11.7	0	7.5	0	GS gLdD_*g*$^+$a	7.0	**0**	4.6	**0**
SS eLaD_*ag*$^-$ *g*$^-$*g*$^-$	7.6	0	5.5	0	GS gLdL_*g*$^-$*g*$^+$	5.2	**0**	2.9	**0.004**
SS eLaD_*g*$^-$*a ag*$^-$	5.7	0	3.6	0.001	SG aLdL_*aa*	15.3	0	11.4	0
SS eLaD_*g*$^-$*g*$^-$ *ag*$^-$	6.1	0	4.0	0	SG aLdL_*ag*$^-$	7.4	0	4.0	0.001
SS eLaD_*g*$^-$*a g*$^+$*g*$^+$	9.9	0	6.9	0	SG aLdL_*g*$^-$*a*	7.4	0	4.3	0.001
SS eLaD_*g*$^-$*g*$^-$ *g*$^+$*g*$^+$	10.4	0	7.5	0	SG aLdL_*g*$^-$*g*$^-$	7.2	0	4.0	0.001
SS eLaD_*g*$^-$*a g*$^-$*a*	11.6	0	5.5	0	SG aLdL_*g*$^-$*g*$^+$	13.3	0	7.3	0
SS eLaD_*g*$^-$*a g*$^-$*g*$^-$	12.3	0	6.2	0	SG aLdL_*g*$^+$*a*	6.4	0	3.4	0.003
SS eLaD_*g*$^-$*g*$^-$ *g*$^-$*a*	11.8	0	5.9	0	SG aLdL_*g*$^+$*g*$^+$	6.5	0	2.9	0.008
SS eLaD_*g*$^-$*g*$^-$ *g*$^-$*g*$^-$	12.7	0	6.9	0	SG gLdL_*ag*$^-$	6.1	0	3.3	0.004
SS eLaD_*g*$^+$*a ag*$^-$	9.9	0	5.6	0	SG eLdL_*ag*$^+$	11.4	0	6.1	0
SS eLaD_*g*$^+$*g*$^+$ *g*$^+$*g*$^+$	5.0	0	4.2	0	SG gLdL_*g*$^-$*a*	7.0	0	4.4	0.001
SS eLaD_*g*$^+$*a g*$^-$*a*	15.2	0	8.0	0	SG gLdL_*g*$^-$*g*$^-$	7.2	0	4.7	0
SS gLaD_*aa ag*$^-$	6.6	0	3.7	0	SG gLdL_*g*$^-$*g*$^+$	11.2	0	5.8	0
SS gLaD_*ag*$^-$ *ag*$^-$	6.0	0	2.8	0.002	SG gLdL_*g*$^+$*g*$^+$	0.0[s]	**1.000**	0.0[t]	**0.981**
SS gLaD_*g*$^-$*a g*$^-$*g*$^+$	11.4	0	6.8	0	GV aLdL_60	1.6	**0.060**	2.6	**0.013**
SS eLaD_*g*$^-$*g*$^-$ *g*$^-$*g*$^+$	11.8	0	7.2	0	GV aLdL_180	0.0[u]	**0.819**	0.0[x]	**0.948**
SS gLaD_*g*$^+$*g*$^+$ *ag*$^-$	0.0[i]	0.951	0.0[j]	0.196	GV aLdL_300	1.1	0.120	1.9	0.039
SS gLaD_*g*$^+$*a g*$^+$*g*$^+$	3.5	0.003	2.9	0.001	GV eLaD_60	4.0	0.001	4.9	0
SS gLaD_*g*$^+$*g*$^+$ *g*$^-$*a*	5.6	0	2.7	0.002	GV eLaD_180	5.3	0	4.7	0
SS gLaD_*g*$^+$*g*$^+$ *g*$^-$*g*$^-$	6.6	0	3.6	0	GV eLaD_300	6.0	0	6.2	0
SS gLdL_*g*$^+$*g*$^+$ *g*$^-$*g*$^+$	2.4	0.018	0.6	0.081					
VG aLdL_60	1.2	**0.080**	1.8	**0.025**	SV aLdL_*g*$^+$*a*_60	1.6	0.031	2.4	0.013
VG aLdL_180	2.3	0.012	2.5	0.008	SV aLdL_*g*$^+$*g*$^+$_60	1.7	0.029	1.8	0.036
VG aLdL_300	1.1	**0.087**	0.7	**0.152**	SV aLdL_*g*$^+$*g*$^-$_60	8.8	0	7.6	0
VG eLdL_60	1.0	**0.108**	0.9	**0.123**	SV aLdL_*g*$^+$*a*_180	1.2	0.065	4.1	0.001
VG gLdL_180	0.0[y]	**0.588**	0.0[z]	**0.521**	SV aLdL_*g*$^+$*g*$^+$_180	1.9	0.019	2.8	0.007
VG eLdL_300	0.9	**0.125**	0.7	**0.172**	SV aLdL_*g*$^+$*a*_300	0.0[ag]	0.491	1.9	0.033
AS aLdL_*ag*$^-$	1.3	**0.090**	0.1	**0.450**	SV aLdL_*g*$^+$*g*$^+$_300	0.4	0.234	0.0[ah]	0.751
AS aLdL_*g*$^+$*g*$^+$	0.0[v]	**0.836**	0.0[w]	**0.504**	SV aLdL_*g*$^+$*g*$^-$_300	5.9	0	6.1	0
AS aLdL_*g*$^-$a	7.2	0	3.6	0.001	SV eLaD_aa_60	9.9	0	10.0	0
AS aLdL_*g*$^-$*g*$^+$	7.3	0	3.4	0.002	SV eLaD_aa_180	5.7	0	8.6	0
AS aLdL_*g*$^-$*g*$^-$	4.9	0	3.1	0.003	SV eLaD_aa_300	8.1	0	9.6	0
AS eLaD_*ag*$^-$	1.4	**0.074**	1.5	**0.038**	SV eLaD_*g*$^-$*a*_60	7.0	0	8.8	0
AS eLaD_*g*$^+$*g*$^+$	6.2	0	5.1	0	SV eLaD_*g*$^-$*g*$^-$_60	7.4	0	9.2	0
AS eLaD_*g*$^-$*a*	7.2	0	3.6	0.001	SV eLaD_*g*$^-$a_300	6.2	0	7.0	0

Table 4 (cont.)

AS eLaD_g^-g^+	7.0	0	4.6	0	SV eLaD_g^-g^-_300	6.6	0	7.5	0
AS eLaD_g^-g^-	8.1	0	4.3	0	SV gLaD_g^-g^-_180	4.8	0	7.3	0
SA aLdL_aa	8.8	0	8.8	0	SV gLaD_g^-a_180	4.5	0	7.0	0
SA aLdL_ag^-	0.9	**0.082**	1.0	0.101	VS aLdL_60_ag^-	1.1	**0.059**	1.3	**0.041**
SA aLdL_g^-a	1.0	**0.072**	1.6	0.038	VS aLdL_60_g^+g^+	0.0[ai]	**0.382**	1.0	**0.062**
SA aLdL_g^-g^-	0.8	**0.100**	1.3	0.063	VS aLdL_60_g^-a	4.3	0	4.9	0
SA aLdL_g^-g^+	6.8	0	4.4	0	VS aLdL_60_g^-g^+	7.1	0	4.5	0
SA aLdL_g^+a	0.0[aa]	**0.388**	0.6	0.219	VS aLdL_60_g^-g^-	1.9	0.015	5.1	0
SA aLdL_g^+g^+	0.1	**0.345**	0.0[ab]	0.579	VS aLdL_180_ag^-	2.2	0.009	1.9	**0.014**
SA gLaD_ag^-	2.8	0.004	4.3	0	VS aLdL_180_g^+g^+	0.7	**0.116**	1.4	**0.033**
SA gLaD_g^-a	2.5	0.006	4.7	0	VS aLdL_180_g^-a	6.3	0	6.0	0
SA gLaD_g^-g^-	2.8	0.003	5.1	0	VS aLdL_180_g^-g^+	8.1	0	5.2	0
AV aLdL_60	1.2	**0.116**	1.8	**0.038**	VS aLdL_180_g^-g^-	4.3	0	6.3	0
AV aLdL_180	1.6	**0.056**	1.1	**0.134**	VS aLdL_300_ag^-	1.1	**0.056**	0.2	**0.244**
AV aLdL_300	0.0[ac]	**0.821**	0.0[ad]	**0.827**	VS aLdL_300 g^+g^+	0.2	**0.293**	0.0[aj]	**0.351**
AV eLaD_60	5.5	0	6.2	0	VS LdL_300_g^-a	4.5	0	0.3	**0.220**
AV gLaD_180	2.8	0.007	4.4	0.001	VS aLdL_300_g^-g^+	7.2	0	3.5	0.001
AV eLaD_300	4.8	0	4.6	0	VS aLdL_300_g^-g^-	2.0	0.012	3.9	0.001
VA aLdL_60	0.1	**0.415**	1.2	**0.115**	VS eLaD_60_ag^-	2.1	0.011	2.6	0.005
VA aLdL_180	1.2	0.060	1.9	0.036	VS eLaD_60_g^+g^+	3.8	0.001	6.9	0
VA aLdL_300	0.0[ae]	**0.462**	0.0[af]	**0.821**	VS eLaD_60_g^-a	5.0	0	5.5	0
VA eLaD_60	2.4	0.008	3.1	0.004	VS eLaD_60_g^-g^+	8.0	0	5.9	0
VA eLaD_180	1.5	0.035	2.4	0.014	VS eLaD_60_g^-g^-	5.8	0	6.2	0
VA eLaD_300	2.3	0.010	2.8	0.007	VS eLaD_180_ag^-	1.5	0.030	1.7	0.019
VA gLaD_180	2.3	0.010	3.6	0.002	VS eLaD_180_g^+g^+	3.1	0.002	5.3	0
SV aLdL_aa_60	10.3	0	10.9	0	VS eLaD_180_g^-a	4.5	0	4.3	0
SV aLdL_aa_180	10.0	0	12.0	0	VS eLaD_180_g^-g^+	6.9	0	4.8	0
SV aLdL_aa_300	8.1	0	10.1	0	VS eLaD_180_g^-g^-	5.4	0	5.0	0
SV aLdL_g^-a_60	2.5	0.007	3.8	0.001	VS eLaD_300_ag^-	2.0	0.012	2.3	0.008
SV aLdL_g^-g^-_60	2.4	0.009	3.5	0.002	VS eLaD_300_g^+g^+	6.6	0	6.2	0
SV aLdL_g^-a_180	3.0	0.003	4.3	0.001	VS eLaD_300_g^-a	7.5	0	4.9	0
SV aLdL_g^-g^-_180	2.9	0.004	1.4	0.074	VS eLaD_300_g^-g^+	7.9	0	5.2	0
SV aLdL_g^-a_300	1.4	0.046	1.9	0.032	VS eLaD_300_g^-g^-	8.4	0	5.5	0
SV aLdL_g^-g^- 300	1.2	0.062	1.6	0.050					

[a] Computed probabilities are for 300 K. [b] Side-chain rotamers of serine (S) are denoted by the variation of g^-, a, and g^+, while those of valine (V) with the use of 60, 180, and 300 (Those of alanine (always g^+) are not indicated). [c] Relative energies (ΔE) are in kcal mol^{-1} relative to the global minima of the given conformational library. [d] All significant probabilities of a conformational library are highlighted (bold). [e] $E_{total} = -579.3093711$. [f] $E_{total} = -586.10480350$. [g] $E_{total} = -656.9586966$. [h] $E_{total} = -664.75725820$. [i] $E_{total} = -805.8411212$. [j] $E_{total} = -815.23882492$. [k] $E_{total} = -812.2403976$. [l] $E_{total} = -822.05155100$. [m] $E_{total} = -618.1341436$. [n] $E_{total} = -625.43014580$. [o] $E_{total} = -618.1357909$. [p] $E_{total} = -625.43265730$. [q] $E_{total} = -692.5752713$. [r] $E_{total} = -700.67186780$. [s] $E_{total} = -692.5804016$. [t] $E_{total} = -700.67614748$. [u] $E_{total} = -695.7754949$. [x] $E_{total} = -704.07756300$. [y] $E_{total} = -695.7762045$. [z] $E_{total} = -704.07945950$. [v] $E_{total} = -731.3997838$. [w] $E_{total} = -739.99859370$. [aa] $E_{total} = -731.3949746$. [ab] $E_{total} = -739.99782330$. [ac] $E_{total} = -734.6000847$. [ad] $E_{total} = -743.40434580$. [ae] $E_{total} = -734.5991759$. [af] $E_{total} = -743.40415690$. [ag] $E_{total} = -809.0370312$. [ah] $E_{total} = -818.64524330$. [ai] $E_{total} = -809.0399540$. [aj] $E_{total} = -818.64600680$.

orientation, in some cases the type II β-turn backbone conformation is shifted from the typical $\varepsilon_L\alpha_D$ into another variant. Most typical conformational shifts are the $\varepsilon_L \Rightarrow \gamma_L$ and the $\alpha_D \Rightarrow \delta_L$, which result in alternative structures that are still part of the type II β-turn family. Thus, type II β-turns have an $\varepsilon_L\alpha_D$ prototype called the "major" form, which occasionally shifts into one of its "minor" forms, such as $\varepsilon_L\delta_L$, $\gamma_L\alpha_D$, and $\gamma_L\delta_L$. Regardless of these minor shifts, all of these structures remain a type II β-turn fold.

In the analysis of secondary structure preferences of proteins, crystallographic data are frequently regarded as the ultimate source of information. The question of how well computed geometrical properties correlate with structural information derived from X-ray data of peptides and proteins is of interest. To find the answer, X-ray results were compared with computed ab initio data both in terms of structure and stability. During such comparisons two questions associated with the analysis of X-ray data of proteins were handled with special care:

1) how to distinguish overlapping structural units (e.g. type I β-turns have a backbone fold similar to that of an α-helix), and
2) what deviation of a particular structure from the "ideal" secondary structure is to be regarded as acceptable?

The total number of β-turns assigned in the experimental database varies from a few up to hundreds of structures (Table 5), depending on the amino acid composition. The assignment of both type I(III) and type II β-turn conformations were based on the degree of similarity between the experimental torsional values and the "ideal" backbone parameters. The ideal type I β-turn has $\phi_{i+1} = -60^\circ+\kappa$, $\psi_{i+1} = -30^\circ+\kappa$, $\phi_{i+2} = -90^\circ+\kappa$, $\psi_{i+2} = 0^\circ+\kappa$ torsional values, while the corresponding data for type II β-turns is $\phi_{i+1} = -60^\circ+\kappa$, $\psi_{i+1} = 120^\circ+\kappa$, $\phi_{i+2} = 80^\circ+\kappa$, $\psi_{i+2} = 0^\circ+\kappa$, as defined by Vancatachalam.[19] These four values (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}) define the center of a 4D-sphere of radius κ ($\kappa = 30^\circ$ or 45°). All dipeptides retrieved from protein X-ray structures with backbone parameters equal to ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}

Table 5. Number of type I(III) and type II β-turns as observed in X-ray determined proteins for all 16 peptide conformation libraries.

	Helices are excluded from the database (without)				Helices are included in the database (with)			
	($\kappa = 30°$[a])		($\kappa = 45°$[a])		($\kappa = 30°$[a])		($\kappa = 45°$[a])	
	type I(III)	type II	type I(III)	type II	type I(III)	type II	type I(III)	type II
GG	11	20	17	28	25	20	79	28
AA	59	1	84	3	212	1	752	4
SS	62	1	91	1	140	1	306	1
VV	7	0	14	0	26	0	154	0
GA	17	3	26	3	46	3	203	3
AG	40	108	56	120	94	118	228	132
GS	28	1	41	2	54	1	108	2
SG	27	64	34	78	44	67	98	82
GV	5	0	7	0	21	0	97	0
VG	19	45	25	58	28	49	111	64
AS	65	1	89	2	151	1	317	2
SA	31	1	48	2	81	1	277	3
AV	16	0	26	0	61	0	313	0
VA	19	3	36	2	48	2	329	2
SV	10	0	13	0	47	0	139	0
VS	29	1	38	2	57	2	169	4
total	445	249	645	301	1135	266	3680	327

[a] Number of turns observed in our protein database. Both type I(III) and type II β-turns were extracted from the experimental database using the following torsional criteria: type I ($\phi_{i+1} = -60° + \kappa$, $\psi_{i+1} = -30° + \kappa$, $\phi_{i+2} = -90° + \kappa$, $\psi_{i+2} + \kappa$) and type II ($\phi_{i+1} = -60° + \kappa$, $\psi_{i+1} = 120° + \kappa$, $\phi_{i+2} = 80° + \kappa$, $\psi_{i+2} = 0° + \kappa$). The four torsional variables (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}) define the center of a hypersphere used for structure assignment, κ controls its radius ($\kappa = 30°$ or $45°$). Both α- and 3_{10}-helices can be excluded (case of "without") or can be included (case of "with") in the experimental database analyzed.

fall in the center of the above-defined 4D-hypersphere, while those deviating "slightly" from these values are located near the center. Such hyperspheres (4D in this case) were used to assign backbone structures similar or identical to type I(III) or type II β-turns in proteins. The backbone parameters of a type I(III) β-turn are close to those of α- and 3_{10}-helices ($\phi_{\alpha\text{-helix}}(i) = -54 \pm 10°$ and $\psi_{\alpha\text{-helix}}(i) = -45 \pm 10°$ while $\phi_{3^{10}\text{-helix}}(i) = -60 \pm 10°$ and $\psi_{3^{10}\text{-helix}}(i) = -30 \pm 10°$). Therefore the result of the secondary structure assignment of hairpin structure depends on whether helical segments are excluded (case of "without") or included (case of "with") in the experimental database. As an example, the case of the -Ala-Ala- dipeptide is reported in Table 8. Using the strongest criterion ($\kappa = 30°$ with all helical structures excluded) a total of 60 turns were assigned, among these 59 had a type I(III) fold. For these 59 structures the average ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2} values and their standard deviations were determined. All four averages are very similar to the ideal values of Vancatachalam.[19] When helical structures are not excluded from the analysis, the total number of type I β-turns increases from 59 to 212 cases for $\kappa = 30°$ and from 84 to 752 cases for $\kappa = 45°$ (Table 8). This significant increase is due to the well known fact that alanine frequently adopts a helix-like conformation and that in terms of torsional angles, helices can be regarded as "adjacent" type I(III) β-turns. Even, when the loosest criterion is considered ($\kappa = 45°$ and both α- and 3_{10}-helices are included) the average values remain close to the ideal ones and the standard deviation increases only for type II β-turns. The analysis of -Gly-Gly- peptides resulted in a picture similar to that for -Ala-Ala-, with two differences:

1) the ratio of type I(III) to type II β-turns is much more balanced, and
2) the inclusion of helical structures from the database only moderately increases the total number of turn structures ($31 \Rightarrow 45$ and $45 \Rightarrow 107$).

The first difference is due to the fact that glycine is the most favored amino acid residue at position $(i+2)$ of a type II β-turn. Thus in the conformational libraries of -Xxx-Gly- peptides, the preference for type II β-turn folds is higher. The second difference between the -Gly-Gly- and -Ala-Ala- sequences is caused by glycine's lack of a special preference for helical conformation. In the case of -Gly-Gly- and -Ala-Ala- peptides, similar analyses were performed for all 16 conformational libraries. When helices are excluded from the database (Table 8 A and B) the increase of κ from $30°$ to $45°$ does not significantly change the ratio of type I(III) over type II β-turns ($\beta_I\,[\%]^{\text{AlaAla}}_{\kappa=30} = 98$ and $\beta_I\,[\%]^{\text{AlaAla}}_{\kappa=45} = 97$ as well as $\beta_I\,[\%]^{\text{GlyGly}}_{\kappa=30} = 35$ and $\beta_I\,[\%]^{\text{GlyGly}}_{\kappa=45} = 38$). On the other hand, the inclusion or exclusion of helical structures from the X-ray database for -Gly-Gly- sequences (Table 8 and Figures 2), using either $\kappa = 30$ or $45°$, does modify the same ratio (from $\beta_I\,[\%]^{\text{GlyGly}}_{\kappa=30} = 35$ to $\beta_I\,[\%]^{\text{GlyGly}}_{\kappa=30} = 56$).

The average backbone conformational values (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}) (Table 8), calculated from the X-ray database both for -Ala-Ala- and -Gly-Gly-, are typical β-turn parameters, and are close to the ideal values predicted by Vancatachalam.[19] When comparing the average backbone values of 105 fully optimized ab initio type I(III) β-turns with the "ideal" parameters, a smaller deviation ($\approx 10°$) is observed for the first and a more significant ($\approx 25°$) for the second amino acid residue. These deviations may partially result from the small basis set applied. A similar difference was previously noticed when only For-Ala-Ala-NH_2 conformers were analyzed. In most optimized type II β-turn conformers both the values of ϕ_{i+2} ($\approx 14°$) and ψ_{i+2} ($\approx 25°$) deviate from their ideal values. The higher standard deviation computed for the latter type of conformation results from consideration of both major ($\varepsilon_L\alpha_D$) and minor ($\gamma_L\alpha_D$) forms of type II β-turns. In general, because of the averaging process, the high deviation of ψ_{i+2} from its "ideal" value ($\psi^{\text{ideal}}_{i+2} = 0°$) in β-turns is known[23]

Table 6. Probabilities at 300 K of type I(III) and type II β-turns as determined by ab initio computations and by statistical analysis of X-ray determined proteins for all 16 peptide libraries.

		Theoretical		Experimental			
				helices are excluded from the datbase		helices are included in the datbase	
peptide	type of β-turn	p_i [%] (RHF)[a]	p_i [%] (DFT)[b]	p_i [%] ($\kappa = 30^\circ$, without[c])	p_i [%] ($\kappa = 45^\circ$, with[c])	p_i [%] ($\kappa = 30^\circ$, without[c])	p_i [%] ($\kappa = 45^\circ$, with[c])
GG	type I(III)	52	34	35	38	56	74
	type II	48	66	65	62	44	26
AA	type I(III)	90	98	98	97	100	99
	type II	10	2	2	3	0	1
SS	type I(III)	98	100	98	99	99	100
	type II	2	0	2	1	1	0
VV	type I(III)	99	100	100	100	100	100
	type II	1	0	0	0	0	0
GA	type I(III)	98	96	85	90	94	99
	type II	2	4	15	10	6	1
AG	type I(III)	12	19	27	32	44	63
	type II	88	81	73	68	56	37
GS	type I(III)	98	94	97	95	98	98
	type II	2	6	3	5	2	2
SG	type I(III)	0	1	30	30	40	54
	type II	100	99	70	70	60	46
GV	type I(III)	100	100	100	100	100	100
	type II	0	0	0	0	0	0
VG	type I(III)	18	18	30	30	36	63
	type II	82	82	70	70	64	37
AS	type I(III)	93	96	98	98	99	99
	type II	7	4	2	2	1	1
SA	type I(III)	99	100	97	96	99	99
	type II	1	0	3	2	1	1
AV	type I(III)	99	100	100	100	100	100
	type II	1	0	0	0	0	0
VA	type I(III)	94	97	86	95	96	99
	type II	6	3	14	5	4	1
SV	type I(III)	100	100	100	100	100	100
	type II	0	0	0	0	0	0
VS	type I(III)	94	97	97	95	97	98
	type II	6	3	3	5	3	2

[a] Relative population in percentage as computed at RHF/3-21G level of theory. [b] Relative population in percentage as computed at B3LYP/6-311++G**//RHF/3-21G level of theory. [c] Relative population in percentage as observed in proteins. Turns were extracted from the protein database using torsional criteria both for type I(III) and for type II β-turns: type I ($\phi_{i+1} = -60^\circ + \kappa$, $\psi_{i+1} = -30^\circ + \kappa$, $\phi_{i+2} = -90^\circ + \kappa$, $\psi_{i+2} = 0^\circ + \kappa$) and type II ($\phi_{i+1} = -60^\circ + \kappa$, $\psi_{i+1} = 120^\circ + \kappa$, $\phi_{i+2} = 80^\circ + \kappa$, $\psi_{i+2} = 0^\circ + \kappa$). The four torsional variables (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}) define the center of a hypersphere used for structure assignment, κ controls its radius ($\kappa = 30^\circ$ or 45°). Both α- and 3_{10}-helices can be excluded (case of "without") or can be included (case of "with") in the experimental database analyzed.

Table 7. Cross-correlation (R^2) values between ab initio computed and by statistically determined type I(III) and type II β-turn probabilities for all 16 peptide libraries.

	Theoretical		Experimental			
			helices are excluded from the data base		helices are included in the data base	
	p_i [%] (RHF)[a]	p_i [%] (DFT)[b]	p_i [%] ($\kappa = 30^\circ$ without[c])	p_i [%] ($\kappa = 45^\circ$ with[c])	p_i [%] ($\kappa = 30^\circ$ without[c])	p_i [%] ($\kappa = 45^\circ$ with[c])
p_i [%] (RHF)[a]	*1.000*	*0.985*	**0.941**	**0.943**	**0.900**	**0.751**
p_i [%] (DFT)[b]		*1.000*	**0.963**	**0.965**	**0.908**	**0.750**
p_i [%] ($\kappa = 30^\circ$ without[c])			1.000	0.993	0.963	0.839
p_i [%] ($\kappa = 45^\circ$ with[c])				1.000	0.976	0.865
p_i [%] ($\kappa = 30^\circ$ without[c])					1.000	0.945
p_i [%] ($\kappa = 45^\circ$ with[c])						1.000

[a] Relative population in percentage as computed at RHF/3-21G level of theory. [b] Relative population in percentage as computed at B3LYP/6-311++G**//RHF/3-21G level of theory. [c] Relative population in percentage as observed in proteins. Turns were extracted from the protein database using torsional criteria both for type I(III) and for type II β-turns: type I ($\phi_{i+1} = -60^\circ + \kappa$, $\psi_{i+1} = -30^\circ + \kappa$, $\phi_{i+2} = -90^\circ + \kappa$, $\psi_{i+2} = 0^\circ + \kappa$) and type II ($\phi_{i+1} = -60^\circ + \kappa$, $\psi_{i+1} = 120^\circ + \kappa$, $\phi_{i+2} = 80^\circ + \kappa$, $\psi_{i+2} = 0^\circ + \kappa$). The four torsional variables (ϕ_{i+1}, ψ_{i+1}, $\phi_i + 2$, and ψ_{i+2}) define the center of a hypersphere used for structure assignment, κ controls its radius ($\kappa = 30^\circ$ or 45°). Both α- and 3_{10}-helices can be excluded (case of "without") or can be included (case of "with") in the experimental database analyzed.

Table 8. Number and conformational parameters of type I(III) and type II β-turns composed form -Ala-Ala- and from -Gly-Gly- amino acid residues as found in our experimental data base of proteins.

		#[a]	[%][b]	ϕ_{i+1}[c]($\sigma_{\phi i+1}$)	ψ_{i+1}($\sigma_{\psi i+1}$)	ϕ_{i+2}($\sigma_{\phi i+2}$)	ψ_{i+2}($\sigma_{\psi i+2}$)
-Ala-Ala-							
$\kappa = 30^\circ$ (without)[d]	all	60					
	type I(III)	59	98	− 60 (9)	− 31 (10)	− 76 (13)	− 14 (11)
	type II	1	2	− 47 (0)	125 (0)	87 (0)	− 8 (0)
$\kappa = 45^\circ$ (without)[d]	all	87					
	type I(III)	84	97	− 62 (12)	− 32 (13)	− 74 (15)	− 19 (13)
	type II	3	3	− 62 (28)	141 (12)	72 (17)	− 4 (29)
$\kappa = 30^\circ$ (with)[d]	all	213					
	type I(III)	212	100	− 63 (8)	− 31 (10)	− 77 (12)	− 14 (13)
	type II	1	0	− 47 (0)	125 (0)	87 (0)	− 8 (0)
$\kappa = 45^\circ$ (with)[d]	all	756					
	type I(III)	752	99	− 63 (8)	− 38 (10)	− 69 (11)	− 31 (15)
	type II	4	1	− 69 (27)	147 (14)	67 (17)	3 (27)
-Gly-Gly-							
$\kappa = 30^\circ$ (without)[d]	all	31					
	type I(III)	11	35	− 59 (12)	− 31 (15)	− 80 (16)	− 7 (14)
	type II	20	65	− 56 (10)	133 (8)	83 (11)	0 (14)
$\kappa = 45^\circ$ (without)[d]	all	45					
	type I(III)	17	38	− 65 (15)	− 31 (18)	− 79 (17)	− 12 (16)
	type II	28	62	− 60 (13)	138 (12)	84 (12)	2 (17)
$\kappa = 30^\circ$ (with)[d]	all	45					
	type I(III)	25	56	− 60 (9)	− 27 (13)	− 85 (16)	− 6 (16)
	type II	20	44	− 56 (10)	133 (8)	83 (11)	0 (14)
$\kappa = 45^\circ$ (with)[d]	all	107					
	type I(III)	79	74	− 62 (10)	− 34 (14)	− 73 (16)	− 25 (19)
	type II	28	26	− 60 (13)	138 (12)	84 (12)	2 (17)

[a] Total number of β-turns assigned for the selected dipeptides (e.g. -Ala-Ala-). [b] Percentage of type I(III) and type II β-turns. [c] Average torsional variables and standard deviations in parenthesis. [d] Turns were extracted from the experimental database using torsional criteria: type I ($\phi_{i+1} = -60^\circ \pm \kappa$, $\psi_{i+1} = -30^\circ \pm \kappa$, $\phi_{i+2} = -90^\circ \pm \kappa$, $\psi_{i+2} = 0^\circ \pm \kappa$) and type II ($\phi_{i+1} = -60^\circ \pm \kappa$, $\psi_{i+1} = 120^\circ \pm \kappa$, $\phi_{i+2} = 80^\circ \pm \kappa$, $\psi_{i+2} = 0^\circ \pm \kappa$). The four torsional variables (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}) define the center of the hypersphere used for structure assignment, with κ radius ($\kappa = 30^\circ$ or 45°). Both α- and 3_{10}-helices can be excluded (case with "without") or included (case of "with") in the database.

therefore a larger tolerance ($\psi^1_{i+2} = 0 \pm 50^\circ$) is allowed during data analysis.

Theoretical energy versus experimental population, conformational preference within a structural library: In the following, the question of how well the computed energies correlate with the probabilities of occurrence of these conformers, derived from an X-ray structure database, is considered. A significant correlation between natural abundance and ab initio computed relative stability implies that the X-ray results, commonly regarded as primary standards, do certify quantum mechanical data. Correlation coefficients (R^2) of such comparison are reported in Table 7. Except for the loosely defined "$\kappa = 45^\circ$ with" case, all R^2 values are higher than 0.9, indicating that it is reasonable to compare computed and experimental probabilities. For example, when relative stabilities determined at the RHF/3-21G level are correlated with experimental data ("$\kappa = 30^\circ$ and without" case) the R^2 value is as high as 0.941. When the same type of experimental probabilities are aligned with the stability data obtained by DFT computations (B3LYP/6-311++G**//RHF/3-21G)) the R^2 value is even higher: 0.963 (Table 7 and Figure 3). Finally, when a single library out of the 16, that of the -Ser-Gly- peptide, is removed, the remaining 15 libraries show a cross-correlation as high as $R^2 = 0.986$. In general, the ratio of type I to type II β-turns was determined both from experimental and from quantum mechanical computed data and significant correlation was observed.

For most amino acid residues (other than Gly and Ala), there are a variety of side-chain conformations, which multiplies the number of conformers to be considered. For example, in the case of the -Ser-Ser- peptide, 24 side-chain variants were computed for type I(III) and for type II β-turns. All of these minima have different relative energies (ΔE) (Table 4) and have a significantly different stability. One can compare probabilities of computed β-turn structures with their experimental counterparts in two ways:

1) either each side-chain rotamer is handled individually, or
2) side-chain rotamers belonging to the same type of backbone conformation are handled commonly. In the latter case, the computed and experimental probabilities of the individual side-chain conformers are summed up and normalized (see Equation (1)].

Often, due to specific side chain backbone interaction, even the fold of the β-turn is modified to some extent. For example, in the case of -Ser-Gly- peptides, the structural shift of the major $\varepsilon_L\alpha_D$ to the minor (typically $\gamma_L\delta_L$) form of the type II β-turn is observed. Nevertheless, all of these minor forms still remain type II β-turn conformations and were included when individual probabilities were accumulated into a single value. The following normalized measure was introduced for both types of β-turns depicted in Equations (1) and (2) for the -Ser-Ser- peptide:

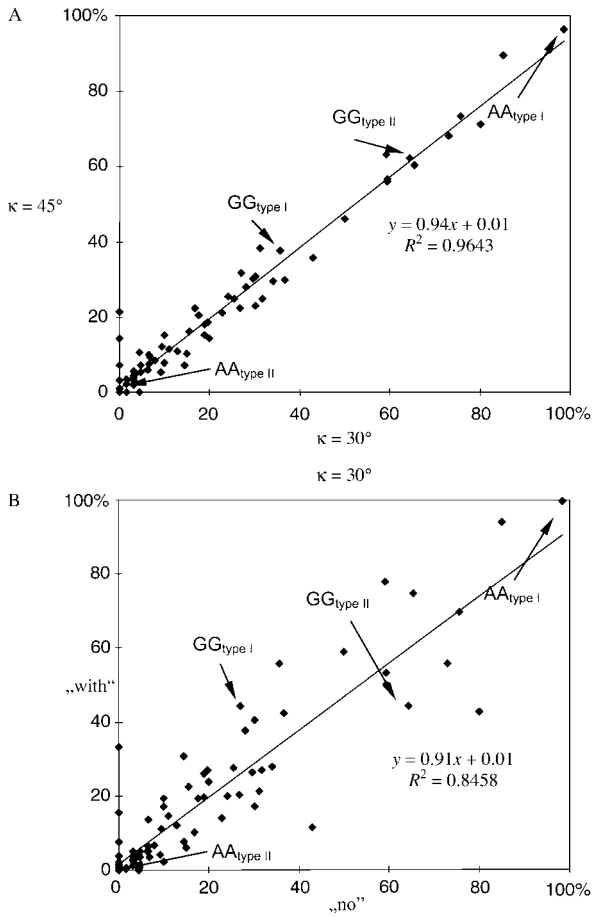


Figure 2. Percentage of type I(III) and type II β-turns as a function of κ and when α- and 3_{10}-helices are excluded ("without", part A) or when included ("with", part B) in the experimental database. Turns were extracted from our experimental database using the following torsional criteria: type I ($\phi_{i+1} = -60^\circ + \kappa$, $\psi_{i+1} = -30^\circ + \kappa$, $\phi_{i+2} = -90^\circ + \kappa$, $\psi_{i+2} = 0^\circ + \kappa$) and type II ($\phi_{i+1} = -60^\circ + \kappa$, $\psi_{i+1} = 120^\circ + \kappa$, $\phi_{i+2} = 80^\circ + \kappa$, $\psi_{i+2} = 0^\circ + \kappa$). The four torsional variables (ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2}, and ψ_{i+2}) define the center of the hypersphere of radius equal to κ. As an example, the percentages of type I(III) and type II conformers associated with -Gly-Gly- and -Ala-Ala- peptides are shown (see Table 9 for more details).

$$p[^{\text{Ser-Ser}}(\text{type I})] = \frac{\sum_{j=1}^{24} p[^{\text{Ser-Ser}}(\text{type I})_j]}{\sum_{j=1}^{24} p[^{\text{Ser-Ser}}(\text{type I})_j] + \sum_{j=1}^{25} p[^{\text{Ser-Ser}}(\text{type II})_j]} \quad (1)$$

$$p[^{\text{Ser-Ser}}(\text{type II})] = \frac{\sum_{j=1}^{25} p[^{\text{Ser-Ser}}(\text{type I})_j]}{\sum_{j=1}^{24} p[^{\text{Ser-Ser}}(\text{type I})_j] + \sum_{j=1}^{25} p[^{\text{Ser-Ser}}(\text{type II})_j]} \quad (2)$$

These cumulative values are markers of the backbone fold (Tables 5 and 6) and vary as a function of the amino acid composition of the peptide moiety. In such a way, theoretical probabilities can easily be compared with the experimental data. Considering these fractions of the -Ser-Ser- model, the computed probabilities at the RHF/3-21G level of theory were 0.98 for type I and 0.02 for type II turns. At a higher level of theory (single-point DFT computations) for the same type of probability of the same peptide, a slightly different number was computed, $p[^{\text{Ser-Ser}}(\text{type I})] = 1.00$ and $p[^{\text{Ser-Ser}}(\text{type II})] = 0$. Experimental probabilities always depend on the actual value of κ, which was set to be either 30 or 45° in this study. The RHF computed $p_i(\text{type-I})/p_i(\text{type-II}) = 0.98/0.02$ (i stands for -SerSer-) ratio matches perfectly with its experimental counterpart p_i("$\kappa = 30^\circ$ without" case) (Table 6). For this peptide the preference for the type of β-turn is determined by ab initio calculations and by statistical analysis of experimental data resulting in identical $p_i(\text{type I})/p_i(\text{type II})$ ratios. This comparison was extended for all 16 peptide models resulting in eight R^2 values of primary importance (Table 7) numbers with bold). The comprehensive analysis of R^2 values shows that the optimum correlation is obtained between single point DFT calculations and $\kappa = 30$ or 45° experimental data (Table 7) and Figure 3). In both cases helical parts were excluded from the experimental database. The R^2 value ≥ 0.96 indicates that the natural abundance of the hairpin conformations of all 16 peptides can be computed with unexpectedly high accuracy.

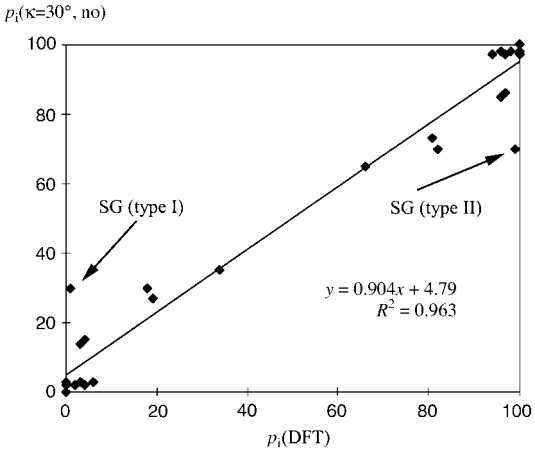


Figure 3. Correlation of B3LYP/6-311++G**//RHF/3-21G (DFT) computed and statistically determined probabilities of type I(III) and type II β-turns for all 16 peptide libraries. (For data see Table 7). Data associated with -SerGly- peptides, the least comparable data, are shown explicitly (see text for more details).

We have shown that the preference for type I(III) over type II β-turns within a conformational library (e.g. -Ser-Ser- or -Val-Ser-) is predicted well by ab initio computations. The correlation is especially good when theoretical probabilities are computed based on single-point DFT calculations. However, it would be interesting to see to what extent these conformational preferences can be compared with each other. Why is it that from the same experimental database a total of 31 -Gly-Gly- and 148 -Ala-Gly- hairpin structures can be extracted? To understand and explain such experimental difference by means of ab initio results, one has to work out how to scale theoretical data into one common frame.

Conformational preference of structural libraries: One of the problems is that in this case the total energies can not be compared directly when all 16 peptides are tabulated in the form of a matrix according to their different amino acid compositions. All "boxes" can be analyzed with respect to

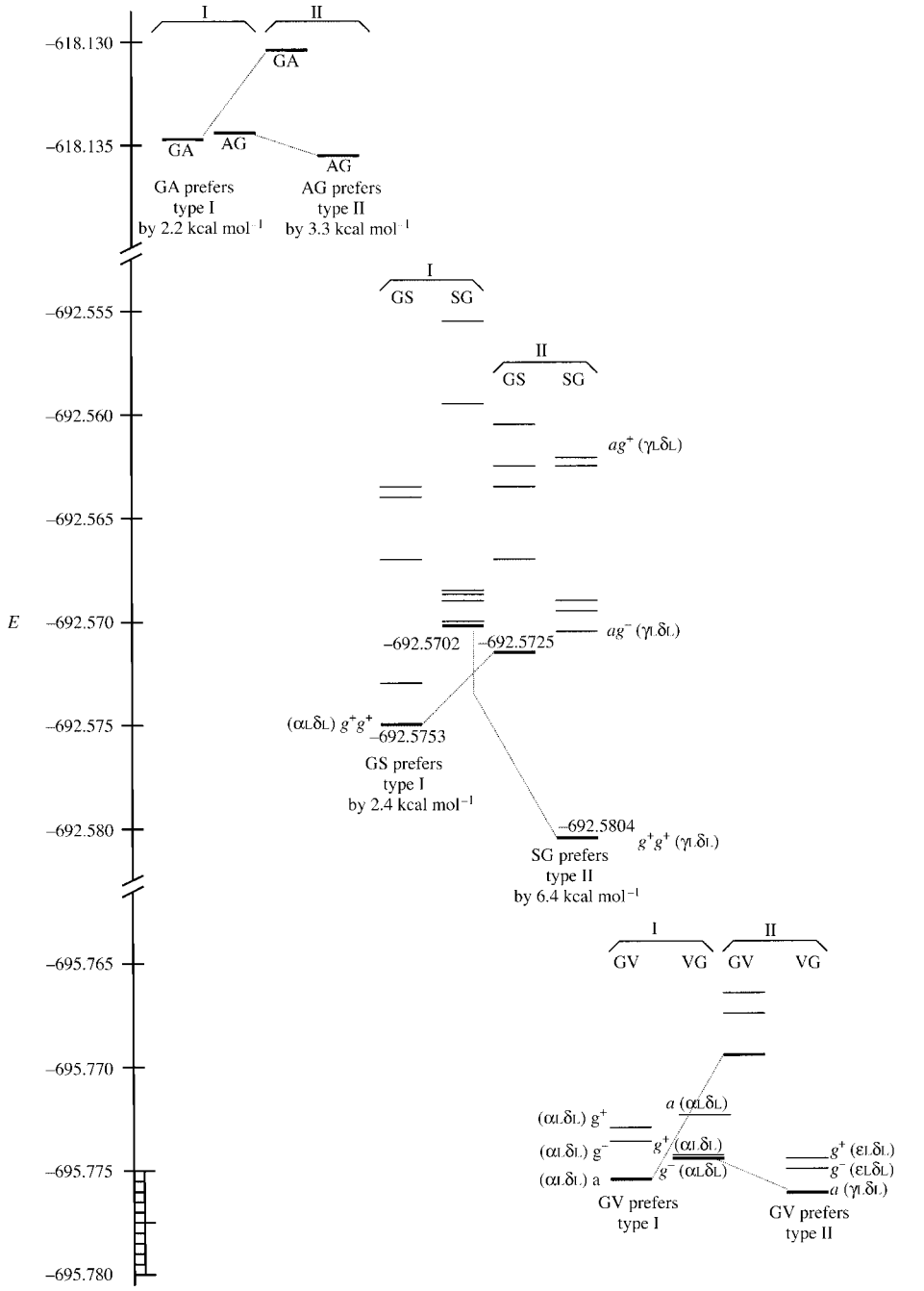


Figure 4. Conformational preference of selected HCO-Xxx-Yyy-NH_2 type models as computed at the RHF/3-21G level of theory. Typically that conformational preference can only be determined within the conformational library.

their preference of type I or type II β-turns individually but not on a common scale. For example, it seems that for GA, GS, and GV peptides the low-energy conformers are all of type I rather than of type II β-turns (Table 4 and Figure 4). Thus, the most stable conformers of AG, SG, and GV peptides are type II β-turns.

One of the other problems we have in comparing total energies is due to the fact that only molecules with identical numbers of the same atoms, that is the isomers, can be compared directly. Thus the hetero subunits -Xxx-Yyy- (e.g. -Gly-Ala-) and -Yyy-Xxx- (e.g. -Ala-Gly-) are comparable but they are neither comparable to the homo subunits -Xxx-Xxx- (e.g. -Gly-Gly-) nor to -Yyy-Yyy- (e.g. -Ala-Ala-). The only obvious way to do this is to compare the average of the total energies computed for the same type of backbone conformation [e.g. type I β-turn, Eq. (3)]. For the above four compounds the averages are identical up to six decimal places. (The difference between the two averages is 8×10^{-7} Hartree at the RHF/3-21G level of theory.)[43]

Such a near equality indicates that the two -CH_3 substituents in -Ala-Ala- exert more or less the same amount of stabilization on the backbone as the two separate -CH_3 groups at the first α-carbon (-Ala-Gly-) and at the second α-carbon (-Gly-Ala-). Thus, the effects of the two methyl groups are practically additive and largely independent of each other. This is, however, a special case and when the backbone or the side chain conformation shifts, or the two substituents interact with each other, such "near-identity" is not expected. In fact the magnitude of the difference [Eq. (4)] may be used as a diagnostic for the extent of such side-chain interactions. The relevant matrices (Figure 5) give such $\Delta\bar{E}$ values in kcal mol^{-1} units for type I and type II β-turns.

$$\bar{E}_{\text{Diagonal}} = \frac{E^{\text{RHF/3-21G}}(\text{Xxx-Xxx}) + E^{\text{RHF/3-21G}}(\text{Yyy-Yyy})}{2} \approx \frac{E^{\text{RHF/3-21G}}(\text{Yyy-Xxx}) + E^{\text{RHF/3-21G}}(\text{Xxx-Yyy})}{2} = \bar{E}_{\text{Off-diagonal}} \quad (3)$$

$$\Delta\bar{E} = \bar{E}_{\text{Off-diagonal}} - \bar{E}_{\text{Diagonal}} \quad (4)$$

When $\Delta\bar{E}$ is *negative* it means that the two off-diagonal isomeric states are stabilized with respect to the two diagonal

isomeric states. In the case of serine- and valine-containing peptides, more than one side-chain conformation was determined, and therefore $\bar{E}$ was computed as the arithmetical average of all side-chain rotamers. This is the case for example for peptides incorporating Gly and Val, since the latter residue may have three side-chain orientations. For example, -Gly-Gly- and -Val-Val- combined are less stable than -Gly-Val- and -Val-Gly- combined for both type I and for type II β-turns ($\Delta\bar{E}^{GV}_{\text{type I}} = -0.09$ kcal mol^{-1} and ($\Delta\bar{E}^{GV}_{\text{type II}} = -0.21$ kcal mol^{-1} at RHF/3-21G). This suggests that there is repulsion between the two *i*Pr groups in Val-Val owing to stereochemical congestion, which is relieved when only one *i*Pr side chain is present in the isomeric diamides: -Gly-Val- and -Val-Gly-. When $\Delta\bar{E}$ is *positive* it means that the two diagonal isomeric states are stabilized with respect to the two off-diagonal isomeric states. This is the case for the Gly and Ser combination for the type I β-turn $\Delta\bar{E}^{GS}_{\text{type I}} = 1.16$ kcal mol^{-1} (and 0.18 kcal mol^{-1} at a higher level of theory). The stabilizing effects of the two hydroxymethyl groups (presumably through hydrogen bonding) are clearly greater in the -Ser-Ser- case than the stabilizing effects of two separate -CH_2-OH side chains, one at the first α-carbon (-Ser-Gly-) and one at the second α-carbon (-Gly-Ser-). Thus, for the same molecules in their type II β-turns, $\Delta\bar{E}$ was found to be smaller $\Delta\bar{E}^{GS}_{\text{type II}} = 0.94$ kcal mol^{-1} (and -0.04 kcal mol^{-1} at a higher level of theory). This implies that when the turn is of type II the two -CH_2-OH groups may stabilize the backbone separately, presumably through backbone–side-chain hydrogen bonding, to a greater extent than was possible for the type I β-turn.

	Gly	Ala	Ser	Val
Gly		0.00	1.16	–0.09
Ala	**0.00**		1.11	–0.08
Ser	**0.18**	**0.23**		0.42
Val	**–0.12**	**–0.22**	**0.23**	

Type I

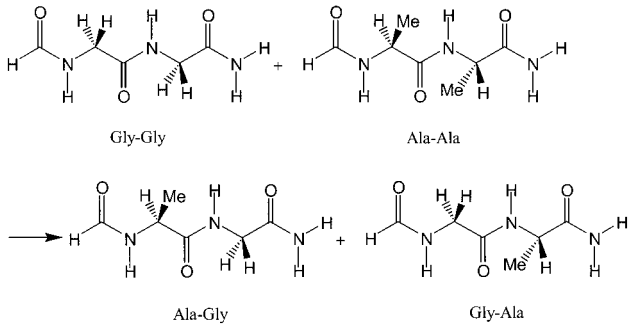
	Gly	Ala	Ser	Val	
		–0.13	0.94	–0.21	Gly
	–0.49		1.19	0.06	Ala
	–0.04	**0.50**		0.75	Ser
	–0.39	**–0.27**	**0.67**		Val

Type II

Figure 5. For both type I(III) and type II β-turns, $\Delta\bar{E}$ values were computed as $\Delta\bar{E} = \bar{E}_{\text{off-diagonal}} - \bar{E}_{\text{diagonal}}$ where $\bar{E}$ is the arithmetical average of all side-chain rotamers. Values determined at RHF/3-21G are in the upper half, those determined at B3LYP/6-311++G**//RHF/3-21G level of theory (bold) are in the lower half of each matrix. (All values are in kcal mol^{-1}.)

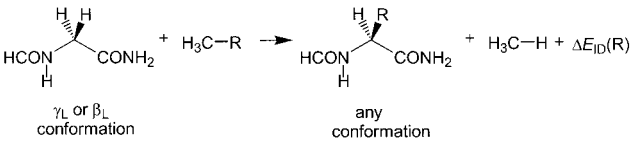
By comparing the computed total energies, we were able to compare nonsymmetric (off-diagonal) sequences such as -Gly-Ala- and -Ala-Gly-. With the averaging technique we were able to relate these off-diagonal elements to their diagonal counterparts, such as -Gly-Gly- and -Ala-Ala-. However we cannot compare all 16 structural families of type I and type II β-turns. To do this we have to construct some isodesmic reactions in order to use the isodesmic energy (ΔE_{ID}) as a comparative energy scale. Of course the choice of reference state predetermines the extent of comparability, but also a single reference state requires more component structures to be optimized.

For example the above averaging technique can be regarded as a simple isodesmic reaction in which one side chain is transferred to a glycine residue (see below) and the isodesmic energy is calculated as $\Delta E_{\text{ID}} = 2\Delta\bar{E}$.



However, this choice of -Gly-Gly- and -Xxx-Yyy- as reference states gives a rather limited scope for comparison. Thus, it is worthwhile to examine other choices for the reference state.

The traditional method involves the replacement of the α-CH_2 group of glycine with the appropriate α-CHR group. This is illustrated below for a single amino acid residue, for the case of a glycine $\Rightarrow$ alanine transformation R = CH_3:



The formula for the corresponding isodesmic energy is shown in Equation (5).

$$\Delta E_{\text{ID}}(\text{R}) = [E(\text{HCONH–CHR–CONH}_2) + E(\text{H}_3\text{C–H})] - [E(\text{HCONH–CH}_2\text{–CNH}_2, \gamma_{\text{L}} \text{ or } \beta_{\text{L}}) + E(\text{H}_3\text{C–R})] \quad (5)$$

For a diamino acid diamide of course two R groups need to be introduced, which may or may not be identical, that is R^1 and R^2.

For this latter reaction the isodesmic energy is calculated as shown in Equation (6).

$$\Delta E_{\text{ID}}(\text{R}^1,\text{R}^2) = [E(\text{R}^1,\text{R}^2, \text{any conf.}) + 2E(\text{CH}_4)] - [E(\text{Gly-Gly}, \gamma_{\text{L}}\gamma_{\text{L}} \text{ or } \beta_{\text{L}}\beta_{\text{L}}) + E(\text{CH}_3\text{R}^1) + E(\text{CH}_3\text{R}^2)] \quad (6)$$

This method is in agreement with previous isodesmic calculations performed on peptides.[44] The energy levels for the above equation using $R^1 = R^2 =$ -CH_3 (i.e., converting -Gly-Gly- to -Ala-Ala-) are shown in Figure 6.

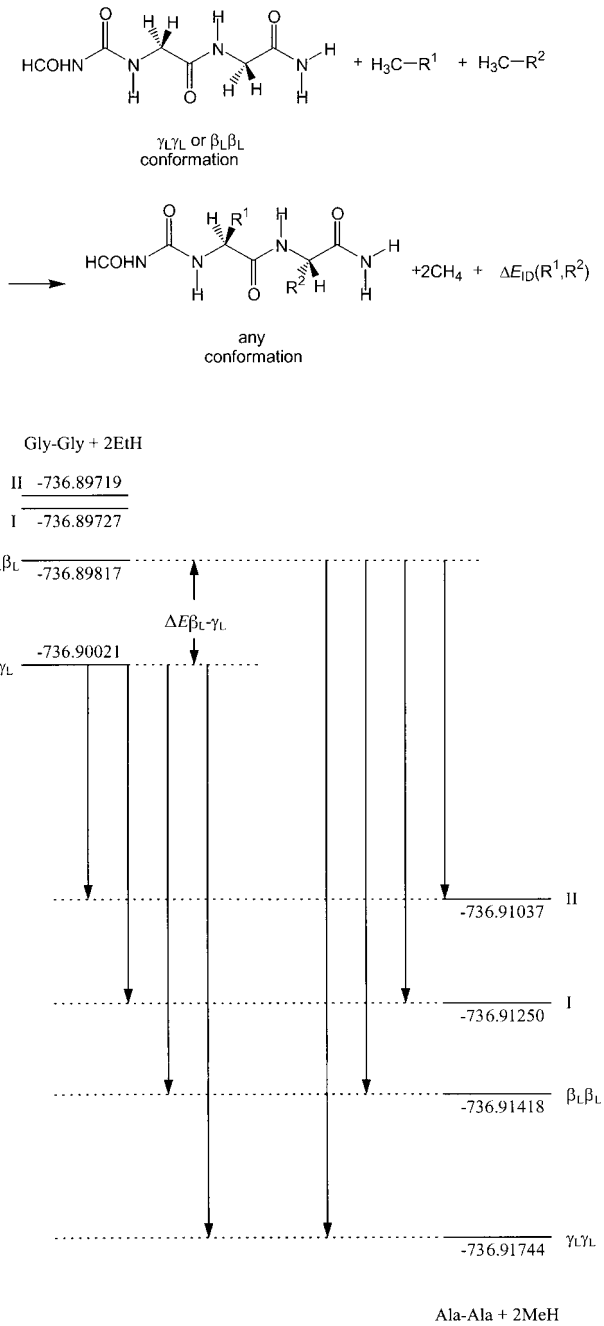


Figure 6. One type of isodesmic calculation scaling any amino acid residue containing peptides (here alanine and glycine) on a common scale.

***β*-Turn type selection**: The original observation that certain amino acid sequences (i.e., primary structures) predetermine the conformation (i.e., secondary structure and overall folding) of a peptide or protein segment is a cornerstone hypothesis in protein chemistry. Based on statistical analysis of X-ray determined protein structures the Chou and Fasman[45] prediction algorithm, together with more recent methods,[46] can predict where β-turns are located along the sequence as well as the probability of adopting either type I(III) or type II forms. One such thesis is that while the -Gly-Xxx- sequence prefers a type I β-turn, -Xxx-Gly- sequences favor type II. This observation has been confirmed by comparison of computed total energies. In analyzing the lowest energy side chain conformers of -Gly-Xxx- and -Xxx-Gly- model systems, in which Xxx = Ala, Ser, or Val, we have found that the experimental rule of β-turn type selection holds, even for these simple model systems (c.f. Figure 4). This suggests that if we anticipate some rule to emerge for the process of β-turn selection, its existence should be based on molecular stabilities.

It should be emphasized again that while -Gly-Ala- and -Ala-Gly- have no distinguishable side-chain conformation, peptides incorporating either Ser or Val have. Consequently, one needs to pay attention to the side-chain orientation in the type I and type II β-turns. As mentioned above in -Gly-Val- and -Val-Gly-, valine may have as many as three, but in -Gly-Ser- and -Ser-Gly- serine may have up to nine side-chain conformers for any given backbone conformer. In Figure 7

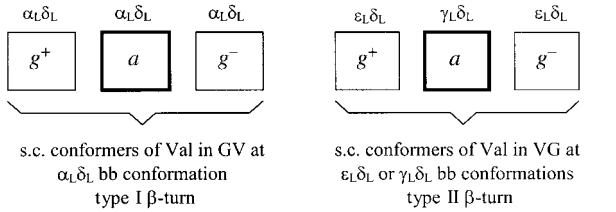


Figure 7. The role of side-chain orientation in Val within the GV and VG triamides in determining type I and type II preference.

below, those side-chain conformers that occur are boxed in and the global minimum is indicated by a heavy-lined square for the hetero (i.e., GV and VG) diamino acid diamide isomer. The analogous information is presented for -Gly-Ser- and -Ser-Gly- β-turns in Figure 8. Only the conformers in a square

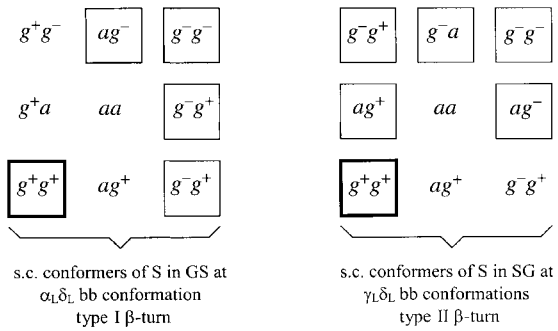


Figure 8. The role of side-chain orientation in Ser within the GS and SG triamides in determining type I and type II preference.

exist and the global minimum is indicated by a heavy-lined square. The energetics of these conformers are illustrated in Figure 4. Taking into consideration all 16 pairs of amino acid residues, it seems that there are only four cases of primary sequences that have certain side-chain conformations in which the type II β-turn is favored over the type I β-turn on energetic grounds, that is in terms of conformational stability. This is illustrated in Figure 9 for the RHF/3-21G computations, where the four cases are AG, SG, VG, and SS. For single point DFT calculations using the larger 6-311++G** basis set the results are slightly different, as shown in Figure 10. At this level of theory the type II β-turn of GG became 0.38 kcal mol^{-1} more stable than type I. Also for the SS case

Table 9. Magnitude of stabilization (average value in kcal mol^{-1}) for the 16 different peptides in their type I and type II β-turn conformation as determined at the B3LYP/6-311++G**/RHF/3-21G level of theory.

	ab initio computed values [kcal mol^{-1}]				
	ΔE_{ID} (average) for type I bb. fold[a]	ΔE_{ID} (average) for type II bb. fold[b]	$\Delta E_{ID}^{\text{type II}} - \Delta E_{ID}^{\text{type I}}$[c]	turn forming potential relative to extended conformation[d]	Experimental turn forming potential as predicted by Chou & Fasman[2],[e]
GG	0.0	0.0	0.0	1.87	**0.016150**[a]
AA	−6.1	−2.8	3.3	−**1.98**	0.002660
SS	−4.0	−2.6	1.4	−**1.77**	**0.017380**
VV	0.0	4.9	4.9	4.15	0.001344
GA	−2.6	0.0	2.7	0.75	0.002975
AG	−3.5	−4.0	−**0.5**	−**1.35**	**0.014440**
GS	−2.8	−0.6	2.2	0.40	**0.010630**
SG	−2.1	−2.5	−**0.4**	−**0.01**	**0.026410**
GV	1.4	5.5	4.1	4.49	0.002380
VG	−0.5	−1.5	−**0.9**	1.01	**0.009120**
AS	−5.9	−3.5	2.5	−**2.22**	**0.009500**
SA	−5.6	−1.9	3.7	−**1.48**	0.004865
AV	−0.8	2.0	2.8	1.35	0.002128
VA	−3.0	−0.2	2.9	0.58	0.001680
SV	−1.1	3.8	4.9	2.46	0.003892
VS	−3.3	−1.1	2.2	0.09	0.006000

[a] Stabilization gained from "side chain". The effects of side-chain interactions are averaged for each peptide individually. Isodesmic values are relative to the type I(III) β-turn conformation of -Gly-Gly-. [b] Stabilization gained from "side chain". The effects of side-chain interactions are averaged for each peptide individually. Isodesmic values are relative to the type II β-turn conformation of -Gly-Gly-. [c] $\Delta E_{ID}^{\text{type II}} - \Delta E_{ID}^{\text{type I}}$. Value smaller than zero indicates the preference of type II rather than type I β-turn structure. [d] Value smaller than zero indicates stabilization of β-turn structure over extended ($\beta_L\beta_L$) conformation. [e] Product of bend frequencies computed on the basis of $(i+1)$ and $(i+2)$ bend frequencies (Table 2A). Value larger than 0.00866 (bold) is associated wit -Xxx-Yyy- peptide as a β-turn former.

$(i+1)$ 1st aa \ $(i+2)$ 2nd aa	G	A	S	V
G	GG **I**	GA I	GS I	GV I
A	AG **II**	AA I	AS I	AV I
S	SG **II**	SA I	SS **II**	SV I
V	VG **II**	VA I	VS I	VV I

Figure 9. Type I and type II β-turn preference among the chosen 16 types of triamide model systems as resulted from RHF/3-21G geometry optimization.

$(i+1)$ 1st aa \ $(i+2)$ 2nd aa	G	A	S	V
G	GG **II**	GA I	GS I	GV I
A	AG **II**	AA I	AS I	AV I
S	SG **II**	SA I	SS I	SV I
V	VG **II**	VA I	VS I	VV I

Figure 10. Type I and type II β-turn preference among the chosen 16 types of triamide model systems as resulted from B3LYP/6-311++G(d,p) single point energy computation.

the type I β-turn became 0.56 kcal mol^{-1} more stable than the type II β-turn.

Unlike AG (−0.5 kcal mol^{-1}), SG (−0.4 kcal mol^{-1}) and VG (−0.9 kcal mol^{-1}), most of these structures prefer type I(III) over a type II β-turn fold (Table 9). This agrees well with experimental observation and predictions: an -Xxx-Gly- sequence in a β-turn does prefer the type II over the type I(III) conformer. However, the preference of these peptides for folding in a β-turn rather than remaining as an extended structure differs. When analyzing the turn-forming potential of these 16 peptides according to theoretical computations (Table 9) the AA, SS, AG, SG, AS, and SA sequences are those stabilized the most and VV, GV, and SV the least. Therefore, peptide units composed from any of the latter three sequences prefer an extended conformation over a hairpin structure. Among these 16 peptides the five sequences that are the most likely to fold in a β-turn are AA (−1.98 kcal mol^{-1}), SS (−1.77 kcal mol^{-1}), AG (−1.35 kcal mol^{-1}), AS (−2.22 kcal mol^{-1}), and SA (−1.48 kcal mol^{-1}). Of these five structures AA, AS, SA, and SS are strong (or medium) type I(III) and AG is a medium strong type II β-turn-forming sequences. The fact that SS, AG, SG, and AS are strong β-turn-forming peptides was recognized some 25 years ago by Chou and Fasman.[45]

Conclusion

Our present study suggests that ab initio determined energies may be of great use in explaining β-turn selection and folding

properties of amino acid sequences, even though solvent effect and long range interactions are ignored. This indicates that some basic rules of peptide folding can be detected and computed even in vacuum and even for short peptides. Such gas-phase computations on small peptides are expected to correlate better with protein X-ray data for segments located in the relatively water-free internal part, than those on the surface of the protein.

Acknowledgements

The research described was supported by grants from the Hungarian Scientific Research Fund (OTKA T032486). The kind help of Anna Füzéry and Ilona Hudáky is appreciated. I.G.C. would like to thank the Ministry of Education (Hungary) for a Szentgyörgyi Albert visiting professorship.

[1] P. Ferrara, A. Caflisch, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10780.
[2] Ashish, A. Grover, R. Kishore, *Eur. J. Biochem.* **2000**, *267*, 1455.
[3] K. Gunasekaran, L. Gomathi, C. Ramakrishnan, J. Chandrasekhar, P. Balaram, *J. Mol. Biol.* **1998**, *284*, 1505.
[4] P. N. Lewis, F. A. Momany, H. A. Scheraga, *Biochem. Biophys. Acta* **1973**, *303*, 211.
[5] P. Y. Chou and G. D. Fasman, *Biochemistry* **1977**, *115*, 135.
[6] a) J. A. Smith, L. G. Pease, *CRC Crit. Rev. Biochem*. **1980**, *8*, 315; b) C. M. Willmot, J. M.Thornton, *Protein Eng.* **1990**, *3*, 479.
[7] A. Perczel, I. Jákli, B. M. Foxman, G. D. Fasman, *Biopolymers* **1996**, *38*, 723.
[8] G. D. Rose, *Nature* **1978**, *272*, 586.
[9] A. Perczel, M. Hollósi, B. M. Foxman, G. D. Fasman, *J. Am. Chem. Soc.* **1991**, *113*, 9772.
[10] G.-B. Liang, C. J. Rito, S. H. Gellman, *J. Am. Chem. Soc.* **1992**, *114*, 4440.
[11] K. L. Constantine, L. Mueller, N. H. Andersen, H. Tong, C. F. Wandler, M. S. Friedrichs, R. E. Bruccoleri, *J. Am. Chem. Soc.* **1995**, *117*, 10841.
[12] A. Perczel, M. Hollósi, P. Sándor, G. D. Fasman, *Int. J. Pept. Protein Res.* **1993**, *41*, 223.
[13] J. Applequist, *Biopolymers* **1981**, *20*, 2311.
[14] J. Applequist, *Biopolymers* **1982**, *21*, 799.
[15] P. L. Polavarapu, Z. Y. Deng, *Faraday Disc.* **1994**, *99*, 151.
[16] G. Jalsovszky, S. Holly, M. Hollósi, *J. Mol. Struct.* (*THEOCHEM*) **1995**, *348*, 329.
[17] K. K. Bhandary, K. D. Kopple, *Acta Crystallogr. Sect. C* **1991**, *47*, 1280.
[18] A. Perczel, G. D. Fasman, *Protein Sci.* **1992**, *1*, 378.
[19] C. M. Vancatachalam, *Biopolymers* **1969**, *6*, 1425.
[20] A. M. Sapse, *Molecular orbital calculations for amino acids and peptides*, Birkhausen, New York, **2000**.
[21] A. Perczel, M. A. McAllister, P. Császár, I. G. Csizmadia, *Can. J. Chem.* **1994**, *72*, 2050.
[22] H. J. Bohm, *J. Am. Chem. Soc.* **1993**, *115*, 6152.
[23] A. Perczel, M. A. McAllister, P. Császár, I. G. Csizmadia, *J. Am. Chem. Soc.* **1993**, *115*, 4849.
[24] M. A. McAllister, A. Perczel, P. Császár, I. G. Csizmadia, *J. Mol. Struct.* (*THEOCHEM*) **1993**, *290*, 181.
[25] J. Frau, J. Donoso, F. Munoz, F. Garcia Blanco, *Biopolymers* **1998**, *45*, 119.
[26] O. Antohi, F. Naider, A-M. Sapse, *J. Mol. Struct.* (*THEOCHEM*) **1996**, *360*, 99.
[27] C. H. Yu, L. Schafer, M. Ramek, *J. Phys. Chem. A* **1999**, *103*, 8337.
[28] K. Möhle, M. Gusmann, H-J. Hofmann, *J. Comput. Chem.* **1997**, *18*, 1415.
[29] K. Möhle, M. Gusmann, A. Rost, R. Cimiraglia, H.-J. Hofmann, *J. Phys. Chem. A* **1997**, *101*, 8571.
[30] R. Vargas, J. Garza, B. P.Hay, D. A. Dixon, *J. Phys. Chem. A* **2002**, *106*, 3213.
[31] R. F. Frey, J. Coffin, S. Q. Newton, M. Ramek, V. K. W. Cheng, F. A. Momany, L. Schafer, *J. Am. Chem. Soc.* **1992**, *114*, 5369.
[32] Based on the analysis of X-ray data of proteins, proline is one of the best candidates for position $(i+1)$ both for type I and for type II β-turns. However, Pro has only one unconstrained backbone torsional angle, ψ. It is too "rigid" and atypical of natural amino acids because of its cyclic side chain. Therefore, instead of proline valine was selected as a hydrophobic amino acid residue.
[33] A. Perczel, Ö. Farkas, I. G. Csizmadia, *J. Am. Chem. Soc.* **1996**, *118*, 7809.
[34] I. Jákli, A. Perczel, Ö. Farkas, A. G. Császár, C. Sosa, I. G. Csizmadia, *J. Comput. Chem* **2000**, *21*, 626.
[35] W. Viviani, J.-L. Rivail, A. Perczel, I. G. Csizmadia, *J. Am. Chem. Soc.* **1993**, *115*, 8321.
[36] I. Jákli, A. Perczel, Ö. Farkas, M. Hollósi, I. G. Csizmadia, *J. Mol. Struct.* (*THEOCHEM*) **1998**, *455*, 303.
[37] G. Endrédi, A. Perczel, Ö. Farkas, M. A. McAllister, G. I. Csonka, J. Ladik, I. G. Csizmadia, *J. Mol. Struct.* (*THEOCHEM*) **1997**, *391*, 15.
[38] A. Perczel, J. G. Ángyán, M. Kajtár, W. Viviani, J.-L. Rivail, J.-F. Marcoccia, I. G. Csizmadia, *J. Am. Chem. Soc.* **1991**, *113*, 6256.
[39] P. A. Karplus, *Protein Sci.* **1996**, *5*, 1406.
[40] Gaussian 98 (Revision A.7), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, Ö. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komáromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian. Inc. Pittsburgh PA, **1998**.
[41] U. Hobohm, M. Charf, R. Schneider, C. Sander, *Protein Sci.* **1992**, *1*, 409.
[42] U. Hobohm, C. Sander, *Protein Sci.* **1994**, *3*, 522.
[43] If Xxx = Gly and Yyy = Ala then at the RHT/3-21G level of theory we obtain the following:

$$-618.1340338 = \frac{-579.3093711 - 656.9586966}{2}$$

$$\frac{-618.1339256 - 618.1341436}{2} = -618.1340346.$$

[44] A. Perczel, P.Hudaky, I. G.Csizmadia, *J. Mol. Struct.* (*THEOCHEM*) **2000**, *500*, 5.
[45] P. Y. Chou, G. D. Fasman, *Biochemistry* **1974**, *13*, 222.
[46] G. D. Fasman in *Prediction of protein structure and the principles of protein conformation* (Ed.: G. D. Fasman), Plenum, New York, **1989**, p. 193–316.

Received: September 5, 2002 [F4393]